JUNE
1951
VOLUME 15
NUMBER 3

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

EDITED BY LAURANCE F. SHAFFER

TEACHERS COLLEGE, COLUMBIA UNIVERSITY

ASSOCIATE EDITORS: EDGAR A. DOLL, *Devereux Schools, Devon, Pa.* WILLIAM A. HUNT, *Northwestern University.* E. LOWELL KELLY, *University of Michigan,* MORRIS KRUGMAN, *Board of Education, New York, N. Y.* BERTHA M. LUCKEY, *Board of Education, Cleveland, Ohio.* FRED MCKINNEY, *University of Missouri.* ANNE ROE, *New York, N. Y.* CARL R. ROGERS, *University of Chicago.* R. NEVITT SANFORD, *University of California.*

MARGARET K. HARLOW, *Managing Editor*

THE JOURNAL OF CONSULTING PSYCHOLOGY is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

EDITORIAL OFFICE

Manuscripts submitted for publication, books and tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y.

Information for contributors: The important characteristics of a good article are the significance of its subject matter, the demonstrated soundness of its conclusions, its clarity, and its brevity. Clarity is improved by careful outlining, discriminating use of headings, and a simple, straightforward style of writing. For brevity, authors should strive to convey every thought without waste of words, and should avoid repetitions of methods, data or findings.

Some of the mechanical requirements for manuscripts are: Type manuscripts with wide margins, on white bond paper. Double space between *each and every line,* including references, footnotes, matter to be set in smaller type, and tables. Use only two types of headings: centered headings, and underlined paragraph headings. Type each table, however small, on a separate sheet, with Arabic number and title. Draw all figures in India ink in final form for reproduction, with draftsman's lettering, but type captions for figures on separate sheets. The *References* must follow the exact form used in APA journals. For further information on the details of preparing manuscripts, see:

> ANDERSON, J. E., AND VALENTINE, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.,* 1944, 14, 345-376.

There is a lag of about nine to twelve months between the acceptance of an article and its publication. Authors may secure immediate publication, however, by arranging for articles to be printed as extra pages, increasing the number of pages that subscribers receive. The cost to the author of such early publication is approximately $11.00 per printed page.

Authors receive fifty reprints gratis, and may order additional quantities when returning proof. Authors are charged one-half of the extra cost of tabular material and cuts, and the full cost of changes made on the proof.

BUSINESS OFFICE

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters should be sent to the American Psychological Association, Inc., 1515 Massachusetts Ave., N. W., Washington 5, D. C.

The Journal of Consulting Psychology is published bi-monthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $5.00, foreign $5.50, single number $1.00.

Copyright 1951 by the American Psychological Association, Inc.

Printed and mailed by the Dentan Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

June, 1951 Vol. 15, No. 3

CONTENTS

Where Are We Going in Clinical Psychology?: CARL R. ROGERS - - - - - - - 171

A Contribution to Rorschach Norms: WILLIAM A. CASS, JR., AND PAUL MCREYNOLDS 178

A Comparison of the Rorschach and Behn-Rorschach Inkblot Tests: ROBERT M. EICHLER - - - - - - - - - - - - - - - - - - - 185

Rorschach Factors Related to Experimental Stress: SYDNEY SMITH AND C. E. GEORGE 190

Color as a Determinant of Responsiveness to Rorschach Cards in Schizophrenia: SAMUEL L. BUKER AND MEYER WILLIAMS - - - - - - - - - - - - - - - - 196

The Ambiguity Values of TAT Cards: SIDNEY W. BIJOU AND DOUGLAS T. KENNY - 203

Factors Affecting the Choice of Certain Szondi Test Pictures: RICHARD WALLEN - - 210

Counseling Method and Client Responses: DANIEL V. BERGMAN - - - - - - - 216

Situational Play Therapy with Normal Children: CLARK E. MOUSTAKAS - - - 225

IQ Variability in Relation to Age on the Revised Stanford-Binet: MURRAY ABORN AND GORDON F. DERNER - - - - - - - - - - - - - - - - - - 231

The Stanford-Binet and the Wechsler Intelligence Scale for Children: ARDEN N. FRANDSEN AND JAY B. HIGGINSON - - - - - - - - - - - 236

The Performance of Congenital Syphilitics on the Wechsler Intelligence Scale for Children: FLORENE M. YOUNG AND VIRGINIA A. PITTS - - - - - - - - - 239

The Relation of Writing Speed to Age and to the Senile Psychoses: JAMES E. BIRREN AND JACK BOTWINICK - - - - - - - - - - - - - - - - - - 243

Internal Consistency of the Group Conformity Rating of the Rosenzweig Picture-Frustration Study: MAHLON V. TAYLOR, JR., AND O. M. TAYLOR - - - - - - 250

MMPI Profiles in Multiple Sclerosis: AARON H. CANTER - - - - - - - - 253

An Index of Adjustment and Values: ROBERT E. BILLS, EDGAR L. VANCE, AND ORISON S. MCLEAN - 257

Item Analysis Aid for the Minnesota Multiphasic Personality Inventory: GEORGE C. CLARK AND ROBERT M. ALLEN - - - - - - - - - - - - - - - - 262

New Books and Tests - 263

WHERE ARE WE GOING IN CLINICAL PSYCHOLOGY?[1]

CARL R. ROGERS
UNIVERSITY OF CHICAGO

FOR twenty-five years I have been deeply immersed in clinical psychology—obtaining my training in this field, engaging in its professional practice, teaching students its science and its art, undertaking research on some of its problems, and taking my part in organizations devoted to its advancement. It has been an interesting and stimulating experience to be so very close to the heart of a young profession and to be both a participant in and an observer of its growing pains. I have been a part of the profession's weak and sometimes sickly childhood; I have been part of its surging adolescence, with its feeling of power and strength, its occasional attitude of infantile dependence, and its opposing and defensive desire for complete independence. It has been a privilege to share in all of these aspects of our growth. I have also participated in the puzzlement of clinical psychology as it faces the future, and endeavors to discover the directions in which it is going. It is to this puzzlement that I wish to devote my paper tonight.

What I would like to do is to portray some of the choices which lie immediately ahead of us, or on which we are already making decisions. I should like to limit my discussion to some of the important issues which will, in all probability, determine the adult character of our adolescent profession. I shall still further limit my consideration to those questions on which we as clinical psychologists differ—either among ourselves or within ourselves. By this I mean problems on which competent clinicians, with the interest of the profession sincerely at heart, take opposing views. Even more, I mean the issues upon which many an individual psychologist finds himself divided, aware of contradictory thoughts and desires within himself. I should like to sharpen and clarify some of these choices by stating them as dichotomies, though they may equally well be thought of as continua. Finally I should like, if possible, to present them as issues for the profession to resolve, rather than to state my own opinion. These are not issues upon which the thinking of any one person is especially important, and to the extent that my own biases show through the material, I shall have failed in my purpose. That purpose is to clarify some of the significant issues on which the whole professional group of clinical psychologists must choose, in their thinking and in their behavior, to take one course or direction, rather than another.

[1]Paper given to the Division of Clinical and Abnormal Psychology, September 4, 1950, as the address of the retiring president of the Division.

PROFESSIONAL DEVELOPMENT OR CREATIVITY?

What are some of these basic choices? The first I will mention might, for the sake of convenience, be labelled the choice between professional development and creativity. Both of these terms represent desirable goals, and we have not always recognized the extent to which they are incompatible. Let us look at the opposing poles of this issue.

On the one hand, our thoughts run to the raising of professional standards. How may we protect the public, offer the maladjusted individual a higher level of service, give public assurance that every clinical psychologist has minimum competence, and protect both the public and the psychologist from

the threat of the quack and the incompetent? We move in various ways to achieve these ends. We set up committees having to do with graduate training to suggest suitable curricula, to advise universities on their present courses and internship programs, to certify those which seem acceptable—in short, to work toward a high-level professional program of preparation in which the basic elements are similar throughout the country, a program which comes up to a minimal standard so far as instructional competence and clinical facilities are concerned.

We also move in other directions to define the profession at a high and satisfactory level. We set up the American Board of Examiners to grant diplomas to those with special competence in clinical psychology. We attempt, through survey and ongoing research, to set up examination procedures which will reliably select the competent clinical psychologist. We work for state certification laws which will give a legal definition to the clinical psychologist. The list of our efforts could be extended much further to indicate the degree to which we are striving to make clinical psychology a profession in which only students with aptitude will be admitted, in which adequate training is adequately standardized, where the degree, the diplomate, and legal certification will all help to define for the public the field in which clinical psychology operates, and those who are qualified to practice it.

But not all of our thoughts run in this direction. There are also advocates of freedom, who point out that the more effective such controls become, the more certain they are to stifle growth and originality and the more surely they will preserve the status quo. They point to the fact that though neither our committees on Graduate Training nor ABEPP have shown a tendency to be coercive, yet their very existence is coercive. There is probably not a university, even at this early date, which would dare to eliminate a course which appears on the recommended curriculm. One of the preoccupations of young psychologists is the speculation as to what the Board examinations will be, and it is the rare individual who will devote a large amount of time to some burning interest of his own which he does not believe will be included in the examination topics. The anxiety over being disapproved hovers over institution and student alike, and has the effect of rewarding conformity and of disapproving originality. We read with new understanding the words which Dr. Alan Gregg addressed to the medical profession some time ago [1, p. 815]:

> It has been somewhat sarcastically observed that military staff-officers prepare for the next war by learning how to fight the last war better. If we devote the present vast energy and attention to raising the standard of performance in what is already known, will American medicine be as alert and adaptable to change as it wisely should? In a world where in some ways change is the only constant, can we be wise to devote our attention away from the ability to change? Are we prepared to believe that nipping the laggards is the essence of leadership, or that enforcing minimal standards will ever discover or encourage originality and advance?

So, as our thinking carries us in this direction of creativity, we ask different questions from those we ask when professional development is our goal. We wish to know: How can we foster originality, variability, and the following of new leads? How can we guarantee that clinical psychology in twenty years will bear little or no resemblance to what it is today? How can we give recognition to fresh thought and creative leadership in our field, whether it comes from those with or without an approved type of training? Can we discover ways of giving recognition for outstanding new achievements in our field, rather than devoting our energies to developing examinations which insure minimal competence? How can clinical psychology preserve one of the most precious traditions of academic psychology—the right of a man to follow anywhere the path that seems to him to lead to truth, and to use the tools most appropriate to that search? How can we achieve that freedom, recognizing with Dr. Gregg that "freedom is more than a pleasure principle: it is a guarantee of vitality and survival through variety. Every tendency in our profession,

especially every trend that seeks to strengthen its position by means of standardization, obligatory uniformity and unvarying acceptance deserves to be challenged as a threat to variety and survival" [1, p. 820].

This choice between professional development and creativity cuts deep, and has much to do with the future of clinical psychology. How much resourcefulness and ingenuity we can bring to it, how much long-range statesmanship, how much ingenious social planning, time will answer.

UNITY OR SPECIALIZATION?

A second basic choice which lies before us may be phrased as a question. Will clinical psychology remain one field, or will it divide into two or more areas of specialization? One line of cleavage might be the separation of the service function from the research function, training some psychologists to be clinical practitioners, and others to carry on the research upon which practice would be based. Another possible basis of division is that the field of psychodiagnosis may develop as one specialty, while psychotherapy becomes a separate field of endeavor. Let us look briefly at some of the considerations which favor unity, and also those which favor division.

Many advantages can be seen in keeping clinical psychology as a broad and comprehensive professional unity. It seems clear that the most fruitful research in personality, psychodynamics, and therapy comes about when the research worker is also a practitioner. Clinical service is the soil out of which significant hypotheses are most likely to grow. Research without roots in practice can all too readily become a sterile laboratory exercise, as we have reason to know.

In regard to the second type of division many questions may be raised concerning the wisdom of dividing diagnosis from therapy. Can one be making an adequate diagnosis if it involves no experientially based concern with treatment? Can one be adequate in his therapy if he is not also thoroughly familiar with diagnosis? Are we not splitting up the client or patient in an unjustifiable fashion?

So the questions and the thinking run. They seem to add up to an impressive array of reasons for keeping the field broad but unified, so that the areas of practice and research, of diagnosis and therapy, may mutually enrich and deepen each other.

Similarly impressive reasons can, however, be marshalled for the opposing point of view. Some see the advance of a science of personality being hampered and confused by the present great demand for psychological service workers. Rosenzweig [3] presents this dilemma clearly and recommends that many psychologists be prepared for a service function while a smaller number be given a parallel but separate training which would point in the direction of research. Others as well have pointed out that since only a small proportion of clinical psychologists develop into productive research workers, we are wastefully deceiving ourselves when it is demanded that all of them should work toward a research degree, the Ph.D.

Considerations of a different sort favor the division between psychodiagnosticians and psychotherapists. It is certainly not possible for any one individual to be thoroughly conversant with all types of diagnostic procedures, and equally impossible (though for different reasons) for one person to be deeply involved in all forms of psychotherapy. It is even more impossible for one clinical psychologist to encompass the totality of both diagnosis and therapy. It is also pointed out that unlike organic medicine, psychotherapy is not to any considerable degree built upon a previously formulated diagnostic study. Still more cogent is the argument that the attitudinal orientation of diagnostician and therapist are decidedly unlike. The personal attitudes which may cause a man to excel in psychological diagnosis may be a severe hindrance to his effectiveness as a therapist, and vice versa. Hence, so the thinking goes—since a different type of personality and a different type of training are needed for psychodiagnosis and psychotherapy, and since the practice of these two

functions is not closely related, why should they not develop as two independent fields of specialization? The recommendation of the Boulder Conference that serious training in psychotherapy be reserved for the postdoctoral level is a step in this direction, though I believe it was motivated by entirely different considerations. Nevertheless, if followed, it means that many psychologists would have training only in the diagnostic function, and it is but an easy step from that to the view that others might have training only in the therapeutic function.

So—unity or division? Which will most enhance the welfare of the persons with whom we deal, and the welfare of clinical psychology itself? It is an exceedingly perplexing issue, one which will be settled piecemeal by every decision we make about the Ph.D., about the psychological service worker at the M.A. level, about the degree of specialization which is permitted to the person working for either degree.

OUR RELATIONSHIP TO PSYCHIATRY

A third issue, one with which we have lived for a number of years, concerns our relationship to medicine and psychiatry. Will our relationship be that of an autonomous, collaborating group, or should our relationship be that of a loyal subordinate group? Or is there some midpoint on this continuum which represents the optimal relationship? Some months ago I had concluded to leave this issue out of my paper since it appeared to be already settled. But recent events indicate that to many individuals, both within and without the profession of psychology, it is an open issue. So I shall endeavor to summarize very briefly considerations which are known to all of you.

On the one hand it is felt that we should be closely linked to medicine and psychiatry in a way which realistically recognizes that society has granted to the medical profession a responsibility for the individual which it has not granted to the psychologist. In order to have a maximal and protected opportunity to function, we should be willing to accept a general supervision from the medical man and to function within medical auspices. This is made the more desirable because psychology and psychiatry have much to contribute to each other. It also gives sound recognition to the unity of the human being as an organism with both biological and psychological aspects.

At the other pole of this issue is the point of view that clinical psychology deals as much with the normal individual as with the person who is ill; with the child in school as with the patient in a hospital; that psychological maladjustments are largely phenomena of learned experience, rather than disease entities; that clinical psychology has an honorable history of an autonomous relationship with several professions, and that this should be continued. It is believed that collaboration with the medical and with other professions is valuable and should be fully developed; that in appropriate situations the psychologist should subordinate himself to the physician, the educator, or the government official, but that such subordination is the decision of a self-responsible professional worker. From this point of view the familar observation is quoted that medicine tends to degrade its ancillary professions, and consequently it is felt that if clinical psychology is to serve its goals of the advancement of knowledge about personality, and the rendering of service to individuals, it is essential that it preserve full professional autonomy.

There can be no doubt that this issue has caused tension and uncertainty in both psychological and psychiatric groups, and any attempt to move toward a particular resolution of the issue seems to meet with a countermovement. For example, the Boulder Conference made a number of conciliatory moves toward a closer linking with psychiatry. This, however, has resulted in considerable feeling that the Boulder Conference "sold out to the medics." On the other hand the joint committee of the psychological and psychiatric associations agreed this spring on a report defining the relationship between the two professions and agreed to present this report to their respective parent bodies. This appeared to be

a long and significant forward step. At the last moment, however, the psychiatric committee failed to present the report to the psychiatric association, evidently with some thought that the report would be rejected as too favorable to the autonomy of the psychological profession. Thus each group vacillates, lacking unity and certainty in its own thinking as to what the optimal relationship should be.

OUR RELATIONSHIP TO THE CLIENT

Another basic choice which we are in the process of making is the decision as to the type of relationship the clinical psychologist is to have to his client or patient. Is that relationship to be primarily the relationship of an expert to a puzzle, or is it to be the relationship of a person to a person? As I observe the situation this is an issue upon which psychologists are deeply ambivalent, though it is an issue which is not often verbalized. Let me try to sketch briefly the two viewpoints which are involved.

On the one hand there is the point of view that it is the application of knowledge which is all-important. The psychologist acquires a knowledge of human beings, just as the chemist acquires a knowledge of chemistry. The major hindrance to advance in our field is the paucity of proven knowledge. The difference between a good clinical psychologist and a poor one is that the first knows much more about personality, about the dynamics of personality, about abilities and aptitudes, and knows more about the application of this knowledge. The individual client or patient is then an intellectual puzzle, an unsolved problem, a chemical unknown, to be analyzed and deciphered by applying to this particular instance the knowledge which has been previously gained. To be sure, this utilization of knowledge, the solution of this crossword puzzle, should be carried on with the thought in mind that this is a human being with whom we are dealing, not a laboratory animal, nor an insect under a microscope. Nevertheless it is the informed intellect of the clinician which is the most significant part of the process. He analyzes, he puts his finger on the crucial elements of difficulty, exactly as the physician does in instances of organic illness. He knows how those crucial elements may be eliminated or modified much as the physician knows how to treat disease. From this point of view any other approach to clinical psychology is sentimental twaddle, which may be tolerated but not respected.

Yet in almost every clinical psychologist there is a vague dissatisfaction with this stand, and it is not surprising that another way of functioning is showing itself. From this angle clinical psychology has to do with the relationship between two individuals as human beings, a relationship largely emotional in character. When an individual is helped, it is perhaps not primarily because of the knowledge which is brought to bear, but because of the wholesome, constructive relationship which has been created. When the concept of clinical psychology runs in these directions, there is concern about the bright, self-assured, beautifully well-informed young men who are earning their degrees in clinical psychology. They have, it is felt, everything except the element which could make them effective, namely, a warm and genuine concern for persons as persons. It is regarded as distressing that psychology bids fair to follow the mistake of medicine in viewing the patient as a complex puzzle rather than as a person first and foremost. It is believed that to see clinical psychology as the skilled application of knowledge is to deny and turn aside from its most potent element—a human relationship between two persons.

It is recognized that this leads into relatively unexplored areas, for there is little tradition or experience in the use of the human capacity for relationships, while there is a wealth of tradition behind the use of the intellect. It is also recognized that to treat another person as a person is to open oneself to change through the influence of the relationship, whereas to deal with him as an unsolved problem is unlikely to demand emotional change in the clinician. Hence it is evident that there is more risk, more uncertainty, in this second approach to clients.

Here then is a profound and cloudy issue which we struggle vaguely to meet and to resolve. Shall we view this individual as a complex and as yet unresolved equation, or shall we enter into a warm, personal, human relationship with him, a relationship meaningful to ourselves as well as to the client? Do we gain our personal satisfactions from the correct understanding and manipulation of a complex set of factors in the case, or from the experience of relating to another individual in a deeply personal way? As a profession, and as individuals, we are not sure where we stand, but are groping in various directions, striving to find the answer.

OUR SOCIAL RESPONSIBILITY

One final puzzlement may be voiced, a final issue upon which we find ourselves ambivalent. Will the clinical psychologist face and consider the social implications of his work, or will he avoid these?

It is hard for us to realize—so rapidly has our profession grown—that what we are doing has social implications of any magnitude. For this reason many psychologists have given no thought to the social consequences of their views because they doubt that there are any significant influences of this sort. They are quite willing to offer their services to any who desire them—to the military for a selection program, to industry for a counseling program, to the VA for passing on pension claims, to the schools for an intelligence testing program, to graduate schools for the selection of clinical psychologists—and they do these things without giving any serious consideration to the meaning which their actions may have for society.

There is another group at this same end of the scale, but for different reasons. In this group we find the psychologist who regards himself as scientist only—who works toward finding and applying the truth which he or his colleagues are able to discover, and leaves to others the interpretation and the understanding of the social significance of what he is doing. He has no concern except to discover and apply the facts and principles which adequately describe human behavior and give the clues to its control. What society chooses to do with these facts and principles is a matter of no urgency for him.

Other clinicians take a different view of their social responsibility. Rosenzweig [4] has discussed the fact that the concept we have of norms has distinct implications for our concepts of man's relation to the state. A stress upon universal norms or upon the statistical average will imply a far more subservient relation to the state than will a stress upon individual norms. Rogers [2] has suggested that the principles which are found to be effective in psychotherapy will, in the long run, be paralleled in patterns of government and philosophies of society. Skinner [5] has endeavored to paint a full-length portrait of society as it would be if some of the concepts of present-day psychology were fully accepted as organizing constructs in the community. The fact that many people—psychologists and others—do not like the picture he paints, only brings more sharply into focus the minimal thinking we have done about our relationship to cultural trends. Hence some psychologists feel that while the effect which our concepts may have upon society should in no way influence us in our search for the truth, we should nevertheless be continually aware of these effects, and should be thinking about them.

Perhaps we can summarize the current state of clinical thinking on this score. Many psychologists do not believe that our concepts have any appreciable influence upon society, and ignore the whole issue. The remainder recognize the problem but adopt differing views. They recognize that whatever concept the psychologist forms as to the nature of man will eventually become woven into the fabric of society. They recognize that the findings, theories, and concepts which the clinical psychologist believes valid in regard to maladjustment and deviations of behavior and personality will eventually find expression in our methods of dealing with criminals, in industry's handling of its employees, in the way children are reared, in the guidance programs of our schools. They realize that the concepts

we form as to the individual's capacity for self-direction, and the extent of his need for external direction, will in the long run be the basis for keeping or discarding democracy, will come to underlie our philosophy of government. But in recognizing these elements, some of us believe that this is a field which we do not enter. We do our work to the best of our ability and let the social consequences fall where they may. Others believe that our responsibility does not cease so quickly. They point to the fact that the atomic scientists, a far more cloistered group, have found it absolutely necessary to develop an interest in the social implications of their studies. They believe that clinical psychologists must participate in clarifying the social meaning of the facts of their budding science, must guard against over-interpretation of these facts, and must provide society with an intelligent basis for choosing its own course of action, for making the value judgments which it must make in regard to the issues posed by psychological science.

CONCLUSION

These, as I see them, are some of the long-range issues which find our profession divided and unsure. How will they be settled?

To some extent they may be settled by discussion. At conferences such as this and even more effectively in regional and state and local groups, in university seminars, in the "bull-sessions" which are an important part of all professional development, some of these problems will be discussed and clarified and conclusions may be reached which will guide behavior. A group will decide that it favors some division of the profession and will actively work toward that goal. A university faculty may decide that it will try to develop a new curriculum which will encourage originality rather than standardization, and will begin to undertake that project. Thus one of the ways in which these issues may be resolved is through discussion by a small group, a group decision, and active moves which implement that decision.

To some small degree they may be settled by official action of the whole profession. The Division of Clinical and Abnormal Psychology might decide that no one would be permitted to call himself a clinical psychologist unless he had completed an approved curriculum and passed a set form of examination. Or a new Division of Psychotherapy might be formed which would in effect divide clinical psychology along one of its cleavage lines. Such official actions, however, unless they are built on the firm decisions of many small groups and individuals, are not likely to be lasting or effective.

Hence we come down to the basic fact that these issues which I have described are settled, in the last analysis, by the day-by-day behavior of each one of us. It is not what I say about the curriculum of clinical psychology but what I do in my classes which is important. It is not what is written about the relationship of psychologist to client which will decide that issue, but the actual relationships which hundreds of psychologists create when they meet new clients or patients next week or next month. Each one of us is operationally deciding each of these issues for himself as he carries on his work, and the sprawling organism which is clinical psychology is affected by each one of these operational choices. It is my hope that if these problems are clarified, if the issues are more clearly seen, that our choices will not be slipshod or ignorant, but will be the outcome of a thoughtful consideration of the deep and conflicting forces which operate in each of us. It is my hope that out of such insightful thinking clinical psychology will move toward becoming a more vital force for the discovery of new truth, and in an increasingly profound sense, a servant of man.

Received September 19, 1950.

REFERENCES

1. GREGG, A. The golden gate of medicine. *Ann. intern. Med.*, 1949, **30**, 810-822.
2. ROGERS, C. R. Divergent trends in methods of improving adjustment. *Harv. educ. Rev.*, 1948, **18**, 209-219.
3. ROSENZWEIG, S. Bifurcation in clinical psychology. *J. Psychol.*, 1950, **29**, 157-164.
4. ROSENZWEIG, S. Norms and the individual in the psychologist's perspective. In M. Reymert (Ed.), *Feelings and emotions: The Mooseheart symposium*. New York: McGraw-Hill, 1950. Pp. 327-335.
5. SKINNER, B. F. *Walden two*. New York: Macmillan, 1948.

A CONTRIBUTION TO RORSCHACH NORMS

WILLIAM A. CASS, JR.
UNIVERSITY OF KANSAS
AND
PAUL McREYNOLDS[1]
VETERANS ADMINISTRATION,[2] PALO ALTO

WE view it as a truism that a test is no better than its norms. Norms do not have to be quantitative, though clearly this is desirable. Neither do norms have to be published and available to all clinicians, though equally clearly this is to be preferred. In the case of the Rorschach test much of the interpretation is based upon subjective norms acquired by the clinician through experience and upon data which are necessarily nonquantitative. Although these facts may well be deplored, they are to a large extent inevitable because of the very nature of the test.

The Rorschach test cannot be considered as altogether a nonquantitative test. On the contrary, we are not aware of any other psychological test which includes *so many* quantitative scoring variables. In view of the wide application of the Rorschach test, it is increasingly difficult to justify the almost complete lack of interest in publishing appropriate normative data for these various scoring categories.

What are the cardinal requirements for adequate Rorschach norms? The following occur to us:

a) The norms should be described in statistical terms. A statistical summary for each category must include at least the mean, the standard deviation, and an indication of the shape of the distribution. The majority of Rorschach scoring categories produce markedly skewed distributions which prohibit their use in statistical analyses requiring assumptions of normality. Further, clinical judgment indicates that in most cases the scale units for a Rorschach scoring category are not equal, e.g., more interpretive weight is given to the difference between two and four *M* responses than to the difference between 22 and 24 *M* responses. These two problems raise the question of whether it might not be profitable to apply normalizing processes to the scoring categories. Converting original Rorschach score values into equivalent *T*-score values of the McCall type [6] would tend to produce comparatively normal distributions for the scoring categories, with approximately equal means and standard deviations. We have not normalized the data of the present study, but are merely suggesting that the process might produce more adequate norms and facilitate research on the Rorschach test.

b) The samples upon which the norms are based should be adequately described. A minimal description would include information as to age, sex, educational level, and occupation. The samples must be drawn in such a manner as to make clear what populations they represent. The populations samples, in turn, should be ones with significance for the Rorschach worker. It is problematical, for example, whether Rorschach norms obtained upon college students furnish an adequate basis for the evaluation of records from the bulk of patients seen in mental hospitals. They may, however, serve admirably as standards in college guidance clinics.

c) There is some question as to whether norms should be based upon normals in the psychiatric sense, or upon normals in the statistical sense. The former possibility would systematically exclude all abnormals, i.e., seriously maladjusted persons, whereas the latter would include simply a random sample from the population in question. Both procedures can be defended; our own inclination is towards the latter.

d) The norms should be in terms of the scoring categories and the scoring system used by the examiner, or to phrase it in a slightly different way, an examiner should use only those norms which are stated in terms of the scoring categories and the scoring system which he employs. Norms based on the Beck system [1], for example, can be used only for certain scoring categories, and then only with extreme caution, by examiners employing the Klopfer and Kelley system [5].

[1]The authors are indebted to Dr. Frances Orr, Stanford University, for facilitating the collection of many of the data used in this study.

[2]Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are the result of their own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

e) The comments in the previous paragraphs refer to objective norms, our major interest in this article. It seems pertinent to suggest, inasmuch as a certain amount of subjective norms is inevitable, that there are rules for the proper accumulation of subjective norms. The examiner should have tested a large number of persons drawn as randomly as possible from the general population. Further, these cases should correspond at least roughly to the patient population with which the examiner is concerned on such criteria as age, sex, educational level, and economic status.

The present status of Rorschach norms leaves much to be desired. To the best of our knowledge, statistical summaries of means, standard deviations, and distribution shapes for the chief scoring categories are not available for the major population groups. Our survey of the literature failed to reveal a single study which reported such minimal statistical data, based on a sizable normal adult sample, for all the major scoring categories of either of the two most widely used Rorschach scoring systems.[8] There is no particular lack in the number of studies bearing upon or discussing norms. The wide variety of scoring systems represented, the dearth of such basic values as standard deviations, the incomplete descriptions of samples studied, and the striking omission of data for many scoring categories make it impossible to combine the results of published material so as to obtain adequate norms.

We offer the hope that in the future Rorschach workers reporting data on normal samples will be particularly careful to include complete statistical data for both the test variables and the samples. Only by following this course of action will it be possible eventually to accumulate sufficient data to guarantee stable and appropriate norms.

Among the most useful statements of norms now available are the summaries presented by Wischner, Rotter, and Gillman [9], and by Brussel and Hitch [2].

The purpose of this paper is to present Rorschach normative data on a sample of 104 normal adult subjects. The collection of cases was made necessary by two other studies [3, 8], and is presented here as a contribution to—rather than as a final statement of—Rorschach norms. It is our hope that this contribution not only will be of practical value to clinicians, but that it will stimulate publication by other workers of normative data in terms which will make possible the combining of various samples so as to increase the stability of norms.

[8]Since this study was prepared for publication, a paper has appeared (S. J. Beck, A. I. Rabin, W. G. Thiesen, H. Molish, and W. N. Thetford, "The Normal Personality as Projected by the Rorschach Test," *J. Psychol.*, 1950, **30**, 241–298) which makes this statement no longer accurate. That paper appears to be an excellent statement of norms according to Beck's system.

PROCEDURE

The sample of normal adult subjects included 58 males and 46 females. Age and education distributions for the sample are presented in Table 1. Some impression of the economic-occupational levels represented may be obtained by listing the occupations reported by the subjects:

TABLE 1

MEANS, SIGMAS, AND PERCENTILE RANK DISTRIBUTIONS BY AGE AND HIGHEST GRADE COMPLETED FOR 104 SUBJECTS

Percentile Rank	Age	Highest Grade Completed
100	75	16
95	67	14
90	61	13
80	53	12
70	44	12
60	39	12
50	34	12
40	29	12
30	27	12
20	25	9
10	22	8
5	20	7
Mean	38.2	11.2
Sigma	14.8	2.4

Baby sitter, baker, bookkeeper (2), business executive (2), butcher, carpenter (4), chauffeur, cleaning woman (2), clerk (6), cook, dishwasher (2), electrician (2), fireman (10), gardener, guard (police) (3), hotel clerk, housewife (21), insurance salesman (2), launderer, laundress, machinist, marine engineer, mechanic (4), merchant seaman, musician, office manager, small store proprietor (2), purchasing agent (retired), salesgirl, salesman (3), secretary (8), service station attendant, show card writer, steam fitter's helper, stenographer, tax investigator, teacher, telephone lineman (2), toolmaker (2), waitress (2), warehouse man, and writer.

No effort was made to classify these occupations by level or type, but they appear to be

fairly representative of the general adult population. None of the subjects was in a professional occupation or attending college at the time of testing, and only two were college graduates. We made no attempt to evaluate the psychological adjustment of the subjects, but it is known that one of them had been discharged recently from a mental hospital. This is probably not too unlike the proportion to be found in the general population.

Thirty-four of the cases were obtained by several well-trained and experienced Rorschach workers who generously contributed their services to the authors. The 70 remaining cases were obtained by members of the graduate class in Rorschach techniques at Stanford University in 1949. Each of the students was required to have administered ten supervised tests before giving any of the tests used in this study. The students were asked to select cases according to certain criteria relating to age, education, sex and economic status. All cases were residents of the San Francisco Bay area at the time of testing, and all were of the Caucasian race. Much credit is due the members of the Rorschach class for exercising remarkable ingenuity and diplomacy in inducing strangers and chance acquaintances to cooperate in taking the test under standard conditions.

All test records were scored according to the Klopfer and Kelley system [5], with the exception of the form level variable, and all scoring was checked by one or the other of the authors. In preparing the scoring category distributions, the magnitude or total score value of a given category for a subject was determined by the sum of all such category scores found in the subject's test record. These sums were obtained by giving a weight of one to each score in the category which was listed as a main score, and a weight of one-half to each score considered as an additional score. The weighting system is not included in the original Klopfer and Kelly procedure, but has found wide usage among practicing clinical psychologists as a means of taking into account response characteristics which are not sufficiently pronounced to be scored as main responses. The weighting system was not applied to those scoring categories which involve percentage ratios (e.g., *F%, W%, A%*), since its use would have distorted the obtained percentage values. Percentage ratios were derived on the basis of the main responses only.

The form level variable was scored according to a variation of the Beck system [1]. Since this is our only deviation from Klopfer and Kelley scoring categories, it merits some explanation. We would have preferred to report *all* categories in Klopfer and Kelley terminology, not because we necessarily consider it superior to the Beck system, but because it appears to be more widely used and because we recognize the need for standardization if other workers are to combine their data with ours. In the case of the form level measure, we judged, correctly or not, that it was best to depart from the Klopfer and Kelley system. It has been our observation that most workers using the Klopfer and Kelley procedure rather carefully evaluate form level, but do not arrive at a quantitative score. The interest here is in obtaining quantitative measures. It is our impression, insofar as a quantitative form level score is concerned, that the Beck system is most widely used. We must assume sole responsibilty for the modification we have made in the Beck procedure. The modification is that all main responses with any *F* component at all (e.g., *cF, FC, M*) were scored for + or —, and were included in the computation of form level. To distinguish this method of scoring from others, we have reported form level estimates as +% rather than *F*+%. The method is more fully described elsewhere [7].

A major problem for the investigator who is summarizing quantitative Rorschach data obtained from a group of subjects is to determine the manner in which the ratio scores should be reported. Cronbach (4) discusses the problem of ratio scores in considerable detail. Not only does Cronbach point out possible arithmetical fallacies in handling ratios statistically, but he notes that derived ratio values are not always in accord with clinical judgment. For example, equal ratios (*W%*) are obtained from two *W* responses in a ten-response record and from 20 *W* responses in a record of 100 responses, yet no clinician would consider the *W* tendencies in the two records to be equal. In the present

study the conventional percentage ratios (e.g., *F%, W%, A%*) were determined for each of the subjects, and a distribution of score values for the sample as a whole, with an appropriate mean and standard deviation, was prepared for each scoring category. We hope that when the normative data reported in this study are used as an aid in interpreting an individual Rorschach record the examiner will evaluate the obtained percentage ratios in terms of the number of responses in the record. Distributions for the other conventional ratios (i.e., *W/M, M/* sum *C, H + A/Hd+Ad*) are presented in the form of two-variable scattergrams. As was indicated previously, it would be somewhat meaningless to determine a single score value for an obtained ratio and then prepare a one-variable distribution with a mean and standard deviation.

TABLE 2

MEANS, SIGMAS, AND PERCENTILE RANK DISTRIBUTIONS FOR THE DETERMINANT SCORING CATEGORIES

Percentile Rank	*M*	*FM*	*m*	*k*	*K*	*FK*	*F*	*Fc*	*c*	*C'*	*FC*	*CF*	*C*
100	20.5	17.5	13.0	6.5	3.0	6.5	39.0	19.0	2.5	5.5	7.5	9.0	3.5
95	8.5	12.0	3.0	2.0	2.5	3.0	23.5	6.0	1.5	4.0	6.0	4.5	1.0
90	6.0	8.0	2.5	1.0	1.5	1.5	17.0	5.0	1.0	3.0	5.0	3.5	0.5
80	4.5	7.0	2.0	1.0	1.0	1.0	13.0	4.0	1.0	2.0	4.0	3.0	0.0
70	3.5	5.5	1.0	0.0	0.5	0.5	10.5	3.0	0.5	1.5	3.0	2.5	0.0
60	3.0	4.5	1.0	0.0	0.0	0.0	9.0	2.5	0.0	1.0	2.0	2.0	0.0
50	2.0	3.5	0.5	0.0	0.0	0.0	7.5	2.0	0.0	1.0	1.5	2.0	0.0
40	1.5	2.5	0.0	0.0	0.0	0.0	7.0	1.5	0.0	0.5	1.0	1.5	0.0
30	1.0	2.0	0.0	0.0	0.0	0.0	6.0	1.0	0.0	0.0	0.5	0.5	0.0
20	1.0	2.0	0.0	0.0	0.0	0.0	4.5	1.0	0.0	0.0	0.0	0.0	0.0
10	0.0	1.5	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	1.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
Mean	3.0	4.5	1.0	0.4	0.2	0.5	9.4	2.6	0.4	1.1	2.1	1.8	0.2
Sigma	3.1	3.3	1.6	1.1	0.8	1.0	6.9	2.5	0.6	1.3	2.0	1.6	0.5

TABLE 3

MEANS, SIGMAS, AND PERCENTILE RANK DISTRIBUTIONS FOR THE LOCATION AND CONTENT SCORING CATEGORIES

	Location									Content				
Percentile Rank	*W*	*W%**	*D*	*D%**	*d*	*d%**	*Dd*	*S*	*Dd + S%**	*A*	*Ad*	*H*	*Hd*	*A + Ad%**
100	28.0	100	51.5	84	16.0	25	29.5	5.5	35	25.5	17.5	24.5	10.5	84
95	18.0	92	27.5	72	6.0	12	9.0	3.5	25	17.5	7.0	7.5	5.5	75
90	16.0	86	21.5	67	3.0	9	6.0	2.0	20	15.5	3.5	6.0	3.5	67
80	12.0	66	17.0	58	1.5	5	3.5	1.5	14	13.5	2.0	5.0	2.0	58
70	10.5	59	15.0	54	1.0	3	2.5	1.0	10	12.0	1.5	3.5	1.5	53
60	9.5	53	12.0	50	0.5	0	2.0	0.5	8	10.5	1.0	3.0	0.5	49
50	9.0	45	9.5	44	0.0	0	1.0	0.5	7	9.0	1.0	2.5	0.5	45
40	8.0	39	8.0	37	0.0	0	1.0	0.5	4	7.5	0.5	2.0	0.5	41
30	7.0	33	6.0	32	0.0	0	0.0	0.0	0	6.0	0.5	1.5	0.5	37
20	6.0	24	4.5	26	0.0	0	0.0	0.0	0	5.0	0.0	1.0	0.0	34
10	4.5	16	2.5	13	0.0	0	0.0	0.0	0	4.0	0.0	0.5	0.0	30
5	4.0	12	2.0	7	0.0	0	0.0	0.0	0	3.0	0.0	0.0	0.0	26
Sigma	9.5	48	11.7	42	1.1	3	2.4	0.9	8	9.4	1.6	3.2	1.2	46
Mean	4.6	24	8.7	19	2.2	5	3.8	1.1	9	4.6	2.7	3.3	2.0	14

*Percentages are based upon the number of main responses only.

TABLE 4

MEANS, SIGMAS, AND PERCENTILE RANK DISTRIBUTIONS FOR THE MISCELLANEOUS SCORING CATEGORIES

Percentile Rank	*R*	*T/R**	*ART-ac*	*ART-c*	*F%**	+%†	*FK + F + Fc%**	*P*	Sum *C*	*NR Last 3 Cds %**
100	87.5	357″	140″	116″	84	100	83	11.0	12.5	75
95	53.0	113	43	47	66	100	73	8.5	7.0	55
90	39.5	78	30	35	59	96	67	7.5	6.0	53
80	31.0	59	21	21	52	93	60	7.0	4.5	48
70	28.0	46	18	19	46	90	54	6.5	4.0	44
60	25.0	40	15	16	41	87	49	6.0	3.5	41
50	22.5	37	13	14	34	84	45	5.5	3.0	39
40	20.0	32	12	12	31	83	42	5.0	2.5	35
30	16.5	26	10	11	27	80	39	4.0	2.0	31
20	14.0	22	8	9	24	79	34	3.5	1.0	29
10	12.0	18	5	6	16	75	27	3.0	0.5	25
5	9.0	15	4	5	9	70	18	2.0	0.0	21
Mean	24.6	46.9	17.2	18.0	37.2	85.0	46.5	5.4	3.1	38.4
Sigma	13.5	47.4	16.6	15.1	16.8	9.3	15.2	2.0	2.1	11.5

*Ratios are based upon the number of main responses only.
†The +% is based upon the scoring system described in the text.

TABLE 5

DISTRIBUTION FOR THE *W/M* RATIO

W	*M* 0-1	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17	18-19	20-21
28-29	1										
26-27											
24-25					1						
22-23											
20-21		1	1								
18-19	1	2		1							
16-17	1	1	1	1					1		
14-15	1		2								
12-13	3	1	2	1							
10-11	5	8	5								
8-9	8	7	5	2	1	1					
6-7	9	8	3	1							1
4-5	6	4	1		2	1					
2-3	1		1	1							
0-1											

TABLE 6

DISTRIBUTION FOR THE *M/Sum C* RATIO

M	Sum C 0-1	2-3	4-5	6-7	8-9	10-11	12-13
20-21		1					
18-19							
16-17			1				
14-15							
12-13							
10-11				2			
8-9	1	1	2				
6-7	3		2	2			
4-5	3	7	7	3	1		
2-3	9	15	7	1			
0-1	7	12	14	1	1		1

RESULTS

In reporting the data we have been desirous of meeting two criteria: (*a*) the data should be presented in a form useful to persons doing research and needing certain normal baselines; and (*b*) the presentation should be of maximal value to the clinician in evaluating a single record. We consider that these criteria have been met by reporting the distributions for the three major ratio scoring categories in terms of scattergrams, and by presenting means, stand-

TABLE 7

DISTRIBUTION FOR THE $H + A/Hd + Ad$ RATIO

	$Hd + Ad$											
$H + A$	0-1	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17	18-19	20-21	22-23
40-43								1	1			
36-39												
32-35												
28-31	1											
24-27									1			
20-23		4		1								
16-19	6	3	5		1			1		1		
12-15	15	10	2	1	1	1	1					1
8-11	15	8	3									
4-7	14	5										
0-3	1											

ard deviations, and percentile rank distributions for each of the remaining major scoring categories.

Normative data based on the sample of 104 adult subjects are given in Tables 2 through 7 for the Rorschach scoring categories and the three score ratios. A representative measure of the Rorschach performance of the sample as a whole can be obtained from a consideration of either the mean or median values for the scoring categories. In view of the skewness of the majority of scoring category distributions, median values are more appropriate estimates of the central tendency of the group of normal adults.

DISCUSSION

The major purpose of this paper is to present data, not to evaluate it. A few remarks of an interpretative nature, however, may be in order. Tables 2, 3, and 4 indicate that the means and medians for most of the scoring categories correspond rather well to the average scores for normals suggested in the literature. There are several exceptions, and these should be pointed out to serve as cautions to examiners seeking "abnormal signs" in Rorschach records. Our data suggest, for example, that *FM* is normally greater than *M*, that *A* is greater than *H*, that *FC* is not greater than *CF*, and that *W*% is equal to or greater than *D*%.

In conclusion, we should like to emphasize that the normative data presented in this paper are not be construed as fixed and ultimate. They represent, rather, the beginning of a statistical standardization which long has been overdue. It may be that future development will imply the need for separate norms for different adult groupings possibly broken down by sex, age, and education.

The normative data herein supplied should not be used mechanically in the clinical situation. In particular, the examiner should keep in mind that much of the richness of Rorschach data is reflected in relationships between scores and in configurations of scores. The norms, with a few obvious exceptions, do not supply standards for evaluating such data. The interpretation of the meaning of various score patterns must, at least for the present, remain within the realm of the clinical intuition of the examiner. It hardly need be added that the "normal" Rorschach psychogram is not necessarily the same as the "optimal" Rorschach psychogram.

Received September 12, 1950.

REFERENCES

1. BECK, S. J. *Rorschach's test. I. Basic processes.* New York: Grune & Stratton, 1944.
2. BRUSSEL, J. A., AND HITCH, K. S. *An introduction to Rorschach psychodiagnostics in military and civilian psychiatry.* Utica, N. Y.: State Hospitals Press, 1947.
3. CASS, W. A., JR. Quantitative Rorschach patterns: A methodological study. Unpublished Doctor's thesis, Stanford Univ., 1950.
4. CRONBACH, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.*, 1949, **46**, 393-429.
5. KLOPFER, B., AND KELLEY, D. M. *The Rorschach technique.* Yonkers, N. Y.; World Book Co., 1946.

6. McCALL, W. A. *Measurement.* New York: Macmillan, 1939.
7. McREYNOLDS, P. Development and evaluation of a new Rorschach technique. Unpublished Doctor's thesis, Stanford Univ., 1949.
8. McREYNOLDS, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.,* 1951, **46,** 131-141.
9. WISCHNER, G. J., ROTTER, J. B., AND GILLMAN, R. D. Projective techniques. In S. W. Bijou (Ed.), *The psychological program in AAF convalescent hospitals,* Army Air Forces Aviation Psychology Program Research Reports, Report No. 15. Washington, D. C.: U. S. Govt. Printing Office, 1947. Pp. 144-157.

A COMPARISON OF THE RORSCHACH AND BEHN-RORSCHACH INKBLOT TESTS[1]

ROBERT M. EICHLER[2]
STATE UNIVERSITY OF IOWA

AT the time Rorschach published his basic text [7], he was quite aware of the necessity for alternate series of inkblots. Thus he indicated:

> Frequently occasion arises when the test must be repeated with the same subject. Such situations appear when one wishes to test normals in various moods, manic-depressives in different stages, schizophrenics in various conditions, or in testing patients before and after psychoanalysis, etc. Or a control test on a normal may be desired. If the test is repeated with the same plates, conscious or unconscious memory enters to warp the result. Analogous series of plates, different from the usual ones but satisfying the prerequisites for the individual plates of the basic series, are necessary for these situations [7, p. 53].

In 1920, Hans Behn-Eschenburg, working under the guidance of Rorschach himself, developed a series of 10 inkblots quite similar in construction to the standard Rorschach test and designed to be equivalent to it. This new series did not become available for general use until 1941, when Zulliger arranged for its publication and wrote a manual for it [12]. Despite the urgent need, both in research and clinical practice, for a parallel series of blots, relatively few investigations in this country have been concerned with determining the comparability of these two tests.

The American literature on the use of the Behn with adult subjects consists of an unpublished thesis and an abstract of a paper presented at an APA meeting. The former study [5] involved giving the Rorschach, Behn, and Harrower inkblots to 30 candidates for psychiatric residency training, the time interval between tests ranging from 6 to 24 hours. One serious limitation of this study was that the brief intertest time interval tended to fatigue the subjects. In addition, the reliabilities for only two determinants were reported. The second investigation [8] compared Behn and Rorschach results obtained on a small and restricted sample of alcoholics. The statistical presentation of the data, however, is too meager to permit adequate evaluation of the degree of correspondence of the two test series.

PROBLEM

The present study was designed to investigate the extent to which the Behn and Rorschach inkblot tests give equivalent results. Specifically, the following questions were raised:

1. With what consistency is a set of measurements made by one test series repeated by the other?
2. Can Rorschach test-retest reliabilities be duplicated by substituting the Behn as initial or final test?
3. What is the comparative frequency of the major scoring categories for both tests?

PROCEDURE

Three groups, each consisting of 35 white male subjects, were tested and retested under the following conditions:

1. Group I. Behn followed by the Rorschach after a median time interval of 20 days (range 14 to 30 days).
2. Group II. Rorschach followed by the Behn after a median time interval of 21 days (range 16 to 23 days).
3. Group III. Rorschach followed by the Rorschach after a median time interval of 21 days (range 18 to 24 days).

[1]The writer is indebted to Dr. Arthur L. Benton for his helpful suggestions during the course of this study.

[2]Public Health Service Postdoctorate Research Fellow of the National Institute of Mental Health. Now at the New York Regional Office of the Veterans Administration.

All but six subjects were enrolled in the introductory psychology course offered at the State University of Iowa. The remaining six were obtained from other courses in psychology and speech pathology. Subjects in Group I were attending the 1949 summer session. Subjects in Groups II and III were obtained during the following 1950 spring semester and were randomly assigned to either of the two conditions. The mean age of subjects in Group I was 25.1 years with a range of 19 to 36 years. The mean age of subjects in Group II was 21.5 years with a range of 20 to 36 years. The mean age of subjects in Group III was 22.1 years with a range of 20 to 32 years. None of the subjects admitted having any experience with the Rorschach or Behn tests. All appeared to be unaware of the specific nature and purpose of the experiment.

Both tests were administered according to the procedure outlined by Beck [1] for administering the Rorschach test, with the single exception that a card was withdrawn after 10 responses had been given to it with the remark, "All right, let us go on to the next one." The test was scored using the scoring symbols of Hertz [3].[3] In computing the scores for each of the symbols, each main response was counted as 1; each additional scoring counted as 0.5. Because of the lack of normative data for the Behn, scores for *Do, Dr, P,* and form quality were not used. For comparability with the Behn, rare details (*Dr*) were considered part of the detail (*D*) score for the Rorschach test.

For each group, the scores for each of the major test factors were subjected to a double classification analysis of variance. The total sum of squares for each factor was analyzed into three components (a) between individuals, (b) between tests, and (c) a residual or error component.[4] The ratio of the mean square for between tests to the error mean square was used to determine the significance of differences between mean scores for each category. This enabled evaluation of practice effect and of the comparative difficulty of the various factors from one test to the other. The relative consistency of measurements for the tests was determined by computation of reliability coefficients. Following the suggestion of McNemar [6], these coefficients were computed by using the ratio of the difference between the mean square for total and the error mean square to the mean square for total, the latter being taken as the best estimate of the trait variation as measured by a single test. This method follows directly from the definition of the reliability coefficient as a function of the error variance relative to the observed trait variance. To determine whether the tests measured with sufficient accuracy to distinguish among the individuals, thus providing a test of the significance of the reliability coefficients, the ratio of the mean square between individuals to the total mean square was used.

Because of the skewed distributions of the test variables, it was necessary to normalize them preparatory to the analysis of variance procedure. To accomplish this the square root transformation was utilized. Where the data tended to be 10 or less, 0.5 was added to the score before computation of the square root [9].

[3] Following Hertz [4] the *Sh wt.* score combines responses to the achromatic colors, gray and black, according to the conventional formula: $FSh = 0.5$, $ShF = 1.0$, $Sh = 1.5$. Because of typographical difficulties, it was impracticable to use the Hertz symbol (M — underlined) for animal movement; the Klopfer symbol *FM* was therefore substituted throughout the paper.

[4] No adjustment was made for the variance due to general productivity.

TABLE 1

RELIABILITY COEFFICIENTS OF SCORES ON THE BEHN AND RORSCHACH TESTS*

($N = 35$ in each group)

	Group I B vs. R	Group II R vs. B	Group III R vs. R
R	.84	.59	.72
W	.52	.66	.82
D	.85	.57	.73
S	.68	.62	.68
M	.51	.50	.82
FM	.51	.46	.46
F	.75	.59	.72
Sh wt.	.80	.51	.55
C wt.	.61	.67	.45
A	.74	.55	.74
H	.54	.56	.76
Anat	.43	.46	.76

*Reliability coefficients obtained by analysis of variance of transformed scores.

RESULTS

Reliability Coefficients. A comparison of the reliability coefficients for the various test scores under the three conditions described above is provided by Table 1. In general, the Behn showed substantial agreement with the measures obtained by the standard Rorschach, with reliabilities for Groups I and II comparing favorably, for the most part, with those obtained under the Rorschach test-retest condition. All coefficients were significant at the 1 per cent level by the test of significance of individual differences. The measures for each of the variables were thus sufficiently accurate to differentiate between individuals.

One might expect on an a priori basis that the test-retest coefficients for Group III would be higher than the coefficients obtained for Groups I or II because of the constancy of sampling of items and the greater influence of the memory factor. However, for only four (*W, M, H, Anat*) of the 12 categories was this true. In Group I the reliabilities for these four categories were the only coefficients failing to equal or exceed the test-retest reliabilities. On the other hand, comparing Groups II and III it is found that the expectation of higher Rorschach test-retest reliabilities was generally realized. Only two of the variables (*FM, C wt.*) for Group II showed coefficients equal to or exceeding those for the Rorschach test-retest condition. Nevertheless, most of the variables in Group II did show a fairly high degree of consistency from Rorschach to Behn in the relative placement of individuals within the group.

The generally higher reliability coefficients found for Group I as compared with Group II can possibly be accounted for by the greater number of responses produced in the former group. The importance of number of responses in affecting reliability was also reflected in the tendency of categories showing small frequencies to be the least reliable. For example, the *Anat* factor, which tended to have one of the smallest frequencies of any of the scoring categories, exhibited the lowest reliability coefficients for Groups I and II, these coefficients also revealing the largest discrepancies on comparison with the coefficient obtained under the Rorschach test-retest condition.

TABLE 2

SIGNIFICANCE (P) OF DIFFERENCES BETWEEN MEAN SCORES ON THE BEHN AND RORSCHACH TESTS* ($N = 35$ in each group)

		Group I			Group II			Group III		
		B	R	P(%)	R	B	P(%)	R	R	P(%)
R	M	30.97	29.37	—	24.40	21.83	—	23.03	26.94	5–10
	SD	17.31	16.76		12.67	8.81		10.74	16.94	
W	M	10.19	6.71	1	7.66	6.09	1	7.40	6.49	1
	SD	3.97	3.90		4.08	4.96		3.16	3.55	
D	M	19.41	21.54	—	16.06	15.20	—	15.50	20.06	2
	SD	15.21	16.13		12.36	8.17		10.04	16.08	
S	M	3.47	2.21	1	1.74	1.56	—	1.21	1.19	—
	SD	2.89	2.90		1.92	1.44		1.14	1.11	
M	M	1.97	2.07	—	2.03	0.99	1	2.37	2.89	—
	SD	1.63	1.63		1.81	0.96		2.52	2.06	
FM	M	4.59	2.73	1	2.11	2.94	1	2.10	2.63	—
	SD	2.58	1.70		1.58	1.47		1.29	1.82	
F	M	12.37	15.91	2	12.94	10.61	5–10	12.63	14.34	—
	SD	9.19	12.01		8.50	5.93		7.40	11.45	
Sh wt.	M	4.71	2.75	1	2.35	2.42	—	1.81	2.59	2
	SD	3.85	2.93		1.94	1.91		1.54	2.38	
C wt.	M	3.91	3.00	1	3.16	2.47	2	2.56	2.52	—
	SD	2.14	1.76		2.29	2.12		1.98	1.57	
A	M	14.00	12.03	1	10.29	10.73	—	9.60	11.01	5
	SD	6.27	6.97		5.81	4.61		4.12	6.21	
H	M	3.83	3.86	—	4.07	3.11	1	4.74	6.37	5
	SD	4.33	4.44		2.45	2.65		4.55	6.79	
Anat	M	1.96	1.54	5–10	2.20	1.27	2	2.21	1.41	1
	SD	1.51	1.56		2.15	1.44		2.06	1.53	

*Significance tests based on F ratios obtained by analysis of variance of transformed scores.

Effects of Repetition. The significance of differences between mean scores for each variable is shown in Table 2. Consistent repetition effects under all three conditions were manifested by the changes in the *W, S, C wt.*, and *Anat* categories. The frequencies for each of these factors declined upon repetition. The decline in *W* was statistically significant in all three groups. *S* decreased significantly in

Group I. *C wt.* decreased significantly in Groups I and II. The decline in *Anat* was significant in Groups II and III and approached statistical significance in Group I.

Effects of repetition were also reflected in the *R* factor. However, these effects were not consistent under all three conditions. While the initial test for Groups I and II tended to produce a larger number of responses, the effect was reversed for Group III, repetition facilitating productivity. Only the latter change approached statistical significance.

Repetition produced several changes in the Rorschach test-retest situation not mentioned above. These include significant increases in *D, Sh wt., A* and *H*. Six of the 12 categories studied showed statistically significant changes, with one additional variable (*R*) revealing a near-significant change.

Effect of Test Form. The Behn lends itself more readily to the production of animal movement (*FM*) responses. The difference between the Behn and Rorschach for this category was significant under both orders of test administration. Tendencies were also shown for the Behn to produce relatively more *Sh wt.* and *A* responses (such differences being significant in Group I) and relatively fewer *D, F, M,* and *H* responses (*M* and *H* differences being significant in Group II, with *F* differences significant in Group I and approximating significance in Group II).

DISCUSSION

It will be noted that the total number of responses to both tests given by subjects in Group I was consistently above the number produced by subjects in the remaining two conditions. Group I also tended to be slightly older than Groups II and III. Groups II and III, on the other hand, were quite similar in mean age and number of responses. In addition, initial treatments were identical for the latter groups, i.e., the standard Rorschach was administered first for these subjects. It would thus seem that the comparison of Groups II and III should be considered as the more crucial one in judging whether Behn results can be employed interchangeably with those secured by means of the Rorschach inkblots. While the Behn generally showed substantial consistency with Rorschach measurements for Group II, none of the reliability coefficients was high enough to warrant considering the test even minimally useful for individual prediction. If a coefficient of at least .70 is insisted on for this purpose, then none of the coefficients for the group attained this level while 8 of the 12 coefficients in Group III exceeded it. In addition, the sizable discrepancies between Groups II and III for several of the reliabilities (e.g., *M, Anat,* and *H*), even when allowance is made for the possibly somewhat elevated reliability coefficients of Group III, indicate that the use of the overall Behn psychogram in lieu of one obtained from the Rorschach would be hazardous in the individual case.

It should be noted, however, that for some factors (*FM, C wt., Sh wt.,* and *S*), the Behn measurements for Group II were at least as reliable as those obtained from the standard Rorschach. Furthermore, the reliabilities obtained in Group I suggest that with more productive records (protocols with a minimum of around 30 responses), the Behn can duplicate satisfactorily the test-retest reliabilities of a considerable number of Rorschach categories, particularly those involving large frequencies.

In general, then, the data indicate that the Behn cannot be considered as an equivalent form of the Rorschach. However, the test appears to be sufficiently similar to warrant its use in many Rorschach research studies. For example, where the influence of the memory factor is to be minimized, groups of subjects could be equated on various categories by means of the Behn prior to the administration of the Rorschach test, thereby avoiding repetition of the latter. Or where an investigation is concerned with the correlation of certain Rorschach signs (e.g., *S, FM, Sh wt., C wt.*) with other measures, it may be expedient in some situations to employ Behn in place of Rorschach frequencies for these variables.

Rorschach Test-Retest Results. While the investigation has been concerned primarily with a comparison of the Behn and Rorschach tests, findings for the Rorschach test-retest condition merit discussion. On the whole the

different scoring categories showed fairly satisfactory reliability coefficients. However, measurements in the case of three important determinants (*C wt.*, *Sh wt.*, and *FM*) were not precise enough for individual prediction.[5] Granted that Rorschach interpretation is ordinarily made in the context of patterns, if the pattern components are unstable, then the pattern itself will tend to be quite fallible. The results for these categories suggest that such variables should be evaluated in a given record with considerable caution.

It will be recalled that 6 of the 12 Rorschach categories investigated showed statistically significant changes on repetition of the test, while the change in a seventh factor approached significance. Such data support Rorschach's contention that repetition serves to "warp" test results and emphasize the extreme care which must be exercised in evaluating changes found on retesting. This is particularly important in view of the proneness of many clinicians to interpret retest changes as reflecting the effects of psychotherapy.

SUMMARY

The present investigation was designed to determine the extent of correspondence of Behn and Rorschach test results. Three groups, each consisting of 35 white male subjects, were tested and retested under three different conditions. The first group received the Behn, followed, after a median time interval of 20 days, by the Rorschach. The second received the tests in reverse order (Rorschach followed by Behn) with a median time interval of 21 days. As a control, subjects in the third group received the Rorschach, followed, after 21 days, by another Rorschach.

Reliability coefficients involving the Behn, while approximating to a considerable degree Rorschach test-retest results for many of the scoring categories, were generally too low for satisfactory use in the individual case. For several of the categories, however, obtained coefficients duplicated Rorschach test-retest reliabilities. Comparison of mean scores on the two tests revealed consistent differences attributable to test series for some response categories.

In general, the results of the study indicated the Behn to be a quite similar but in no sense parallel test to the Rorschach. Possible important uses of the Behn were discussed. In addition, reference was made to implications of the incidental findings on Rorschach test-retest results.

Received August 7, 1950.

[5]Reliabilities for Rorschach scores determined by previous workers on similar samples are not comparable to the coefficients reported here. Most of the studies [10, 11] employed the split-half technique. One study [2] was concerned with test-retest reliability but utilized an unsound method for statistical analysis of the data.

REFERENCES

1. BECK, S. J. *Rorschach's test. I. Basic processes.* New York: Grune & Stratton, 1944.
2. FOSBERG, I. A. An experimental study of the reliability of the Rorschach technique. *Rorschach Res. Exch.*, 1941, **5**, 72-84.
3. HERTZ, MARGUERITE R. The scoring of the Rorschach ink-blot method as developed by the Brush Foundation. *Rorschach Res. Exch.*, 1942, **6**, 16-27.
4. HERTZ, MARGUERITE R. Suicidal configurations in Rorschach records. *Rorschach Res. Exch.*, 1948, **12**, 3-58.
5. MAYMAN, M. A comparative study of the Rorschach, Harrower, and Behn-Eschenburg ink-blot tests. Unpublished Master's thesis, New York Univ., 1947.
6. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.
7. RORSCHACH, H. *Psychodiagnostics.* (2nd Ed.) New York: Grune and Stratton, 1942.
8. ROSENWALD, A. K. A comparison of the Rorschach and Behn-Rorschach tests based on a study of chronic alcoholic subjects. *Amer. Psychologist*, 1947, **2**, 270. (Abstract.)
9. SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Iowa State College Press, 1946.
10. THORNTON, G. R., AND GUILFORD, J. P. The reliability and meaning of Erlebnistypus scores in the Rorschach test. *J. abnorm soc. Psychol.*, 1936, **31**, 324-330.
11. VERNON, P. E. The Rorschach ink-blot test. *Brit. J. med. Psychol.*, 1933, **13**, 179-205.
12. ZULLIGER, H. *Einführung in den Behn Rorschach test.* Bern: Hans Huber, 1941.

RORSCHACH FACTORS RELATED TO EXPERIMENTAL STRESS[1]

SYDNEY SMITH AND C. E. GEORGE
ARIZONA STATE COLLEGE

SINCE the end of World War II an extensive literature on projective techniques has been presented to psychologists. Singularly lacking in most of these published reports are studies on validation procedures, especially studies based on direct experimentation. Experimental psychology has not been slow in criticising the clinician (who so extensively depends on projective techniques) for not taking the problem of validation more seriously. The clinical use of the Rorschach, as an example, provides many instances of an acceptance of authority and a dependence on untested assumption that are most unbecoming to a scientific subject matter still striving for status among its parent sciences. Some recent writers have been aware of the weaknesses in the projective techniques and have not only indicated the need for validation studies, but have outlined the methods by which such studies might be accomplished [6, 11]. But interest continues to lie in the direction of discovering patterns of test factors for selected populations without always knowing what such factors are actually measuring.

The present study is an attempt to discover by means of direct experimentation using the group Rorschach technique what particular Rorschach factors are related to experimental stress. The general experimental design utilized by the present writers closely parallels that used in a similar study reported by Williams [14] and is one which might be extended to the study of the other test factors. Williams felt the need for explaining to his readers a point that might be briefly reiterated here: Studying individual factors on the Rorschach test is not violating the clinical ideal of measuring the personality as a whole, nor does it alter the usefulness of the total test record where the various factors are known to be to a degree interdependent.[2] It is legitimate procedure to discover, if possible, what psychological meaning a specific Rorschach factor may have. The experiments described here focused attention on the Multiple Choice Group Rorschach technique as developed by Harrower-Erickson [5]. The scoring procedure has been somewhat modified from that used originally by Harrower-Erickson in order to make it similar to the Klopfer [7] system of scoring the individually administered test. Many of the questions that have already been answered experimentally for the individually administered test need to be rephrased and retested in the case of the group test, for one is not safe in assuming that the two instruments, different as they are in their setting and administration, are measuring the same thing in the same way. The group test is a separate instrument and raises questions of control and procedure that need not trouble the examiner in the case of the individual test. Offering evidence for the validity of the group Rorschach test is a large order, and the research presented here explores only a small part of this much larger problem. Specifically, the purpose of this paper is to discover the relationship to certain Rorschach test factors of the changes taking place in test performance on a digit-symbol test and an information test as the result of experimentally induced stress.

EXPERIMENTAL DESIGN WITH THE DIGIT-SYMBOL TEST

General Design and Procedure. Glass slides of the Rorschach plates were shown one at a time in the orthodox sequence to a group of applied psychology students. The plates were pro-

[1]The authors are greatly indebted to Dr. Herbert Gurnee for many helpful suggestions.

[2]Cronbach [2] has recently developed a technique for combining various test factors statistically that will be of value to future researchers.

jected upon a screen, and the subjects were permitted approximately a minute to a minute and a half to view each plate while the room was darkened. At the end of this time the lights were turned on so that the subjects could see the printed Harrower-Erickson Multiple Choice Rorschach booklets on which they were to record their responses. During this interval each plate remained on the screen until everyone was ready for the next one to be shown. Later each subject was asked to make an individual appointment with one of the writers for a continuation of the experiment. When the subject appeared for his appointment, he was asked to sit opposite the examiner at a desk and was handed a blank copy of a prepared digit-symbol test. The directions used were those already standardized for the Wechsler-Bellevue Intelligence Scale [12], but the actual design of the test differed in form and difficulty from that used on the Bellevue Scale. The Digit Symbol Test appearing on the Wechsler-Bellevue Intelligence Scale was thought to be too simple and too easily memorized for the purposes of this experiment. Therefore, the test was not only lengthened but the symbols were made more complex and thus more difficult to duplicate. The subject was permitted four minutes to complete the test, and at the end of that time, the examiner looked over the results making the following comments:

> You didn't do as well on this as I expected. I can see that a number of the symbols caused you real trouble, and that is an indication that you weren't paying attention to what you were doing. Furthermore, you didn't get very far in the time you were alloted, and you ought to be able to improve yourself on that score. I'm going to give you another chance to see if you can't get more of the work done in less sloppy fashion.

At this point the subject was given the test a second time and allowed to work under the same time limit. It was hoped that the criticism would serve as an adequate stress situation, especially since it was coming from the student's class instructor.

Subjects. The subjects comprised 41 members of a class in applied psychology. The age range was from 18 to 43 years, with a mean age of 24 and a standard deviation of 5.19.

EXPERIMENTAL DESIGN WITH THE INFORMATION TESTS

General Design and Procedure. The Rorschach was administered to two additional groups of subjects according to the same procedure as above. The Information tests from Forms I and II of the Wechsler-Bellevue Intelligence Scale [22, 23] were then used for the second half of this part of the experiment. The two information tests had been mimeographed separately, with ample space for recording answers. First, Form I was distributed among the subjects, and at a signal from the examiner, the subjects began working on the test for a period of three minutes. At the end of this time the tests were collected and the examiner made the following comments:

> The results of this test are somewhat disappointing. I can see by glancing through the sheets that most of you didn't complete the test items. All of you should have been able to reach the last questions in the time given you. You ought to be able to do better. This test is taken directly from an adult intelligence test and is a very accurate indication of your intellectual efficiency. In other words this test provides an excellent indication of your IQ. Now here is a second test we are going to take in the same way. I expect you to do much better with these questions since they are designed primarily for children and are really much easier.

Form II was then administered using the same time limitation.

It was believed that the foregoing comments would create a stress situation by informing the subjects that the tests are essentially measures of intelligence and by indicating that the second test is more elementary than the first whereas, in reality, they are of approximately equal difficulty. A correlation coefficient between the two forms is not reported by Wechsler, but numerous other workers have concerned themselves with this problem, including the present writers. Cronbach [1] discovered a correlation of .86 between the two forms of the Information Test, a figure as high as that obtained for test-retest on the same form [4]. The present writers, using alternated presentation of the forms on 33 subjects, obtained an r of .85.

Subjects. Two groups of subjects differing considerably in age range were used. The first group contained 34 subjects in an introductory psychology class, while the second contained 64

people enrolled in an educational psychology class. For convenience the group of applied students mentioned earlier can be referred to as Group I, and the two present classes as Group II and Group III, respectively. Group II displayed an age range of from 17 to 36 years, with a mean age of 22 and a standard deviation of 4.46. Group III had an age range of from 17 to 54 years, with a mean age of 26 and a standard deviation of 9.76.

Scoring. Wechsler's established scoring criteria were used in connection with the Information Test. Two scores were thus obtained for each subject: first, a score on Form I prior to the onset of stress, and second, a score obtained on Form II after the onset of stress. The difference between these two scores served as a measure of the influence of stress on the subject or what can properly be referred to as the behavior criterion of stress. These were then converted to *T*-scores. The digit-symbol tests presented a more difficult problem in scoring. Since the same test was used both before and after the onset of stress, there was a considerable practice effect apparent. Only two of the subjects showed a loss in score on the second test. By adding a constant to each score equal to the greatest loss in score on the second test, the two negative scores were eliminated. For convenience of handling as well as for ease of understanding, the differences between the two digit-symbol tests were converted into *T*-scores, thus making them directly comparable to the differences between the Information tests and permitting the handling of the data from both sets of tests in the same manner.

TABLE 1

RELATIONSHIP BETWEEN THE BEHAVIOR CRITERION AND RORSCHACH FACTORS

	N	*eta's*	*F*	*P*
Form and Behavior				
		.869	6.18	.001
Group II	34			
		.600	1.83	.10
		.450	1.84	.05
Restricted Age Group	108			
		.403	1.90	.05
M/M + FM and Behavior				
		.504	1.08	–
Group II	34			
		.336	0.50	–
		.441	0.90	–
Restricted Age Group	108			
		.298	0.61	–

RESULTS

Table 1 summarizes the relationship between *F%* and the difference scores on the Information and digit-symbol tests. This relationship turned out to be the most significant of those studied, and its curvilinearity is of considerable theoretical interest. The discovery of the true nature of the relationship was made first with Group II, which was the earliest to be dealt with experimentally in point of time. As with most studies the present writers set up the hypothesis that the relationship would be classically linear, but the following figures indicate that the null hypothesis to the effect that the relationship does not depart significantly from linearity can be rejected at the 1 per cent level of confidence. The Pearson product-moment coefficient of correlation equaled .508. The variance ratio (*F*) test of significance was 11.13, which gives a probability of .01. The correlation ratio (*eta*) obtained was .869, with a variance ratio or F_1 of 4.48. The second variance ratio (F_1) measures the significance of curvilinearity. For clarity these figures are summarized in Table 2.

TABLE 2

TEST FOR CURVILINEARITY BETWEEN *F%* AND THE BEHAVIOR CRITERION ON GROUP II

r	*F*	*P*	*eta*	F_1	*P*
.508	11.13	.01	.869	4.48	.01

On the basis of these figures it was decided to test the significance of the relationship under study for the entire sample by combining the three groups of subjects. But in Group III there was basis for concern over the large age deviation, and it was felt necessary to determine the influence of the age factor as well. A scattergram showing the relationship between age and *F%* revealed that subjects of 30 years and over used remarkably fewer *F*'s than younger subjects. A chi-square test ($\chi^2 = 7.80$) indicated that the difference was significant at the 1 per cent level of confidence. For the older subjects, there was not enough range in the *F* scores to justify the use of this criterion for the prediction of intellectual control

under stress.[3] For this reason further study was restricted to that portion of the sample falling between the ages of 17 and 29 years ($N = 108$). The obtained relationship is reproduced in Figure 1, where it can be seen that the area of greatest predictability lies between an $F\%$ of 30 and 50. It is important to point

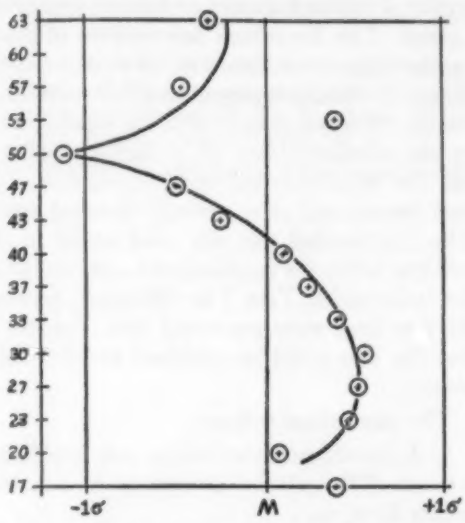

FIG. 1. Rorschach form-stress curve. The vertical axis represents $F\%$ and the horizontal axis represents behavior scores.

out in this connection that the number of Rorschach responses was held constant for every subject—a function of the Multiple Choice Group Rorschach—which solves the problem of reliability for the F per cents.[4] The relationship between form and behavior is reported in Table 1, where both correlation ratios are given. For a discussion of the use of the correlation ratio, see the reference on McNemar [8].

In relation to the problem of stress, the Rorschach factors M and FM were also studied. Klopfer believes the M to FM relationship is an indication of maturity, and the possibility that maturity as measured in this way might be related to intellectual control under stress seemed worthy of investigation. A biserial correlation was run between the behavior scores and the dichotomy of $M \leqq FM$ and $M > FM$. The result was in the expected direction (.33), but too low to be significant. Due to the previous finding of curvilinearity between the behavior criterion and the F factor, it was realized that the low biserial r could not be accepted as an indication of lack of significance until the assumption of curvilinearity had also been tested. Such a test calls for some means of quantifying the M to FM relationship, and Monroe [9] conveniently provides a ratio for this purpose: $M/M + FM$. The authors are fully aware of the undesirability of such ratios, but no other method presented itself. The Monroe ratio was used, and the resulting *eta*'s are reported in Table 1. The difficulties of quantification may be seriously interfering with the discovery of significance here.

DISCUSSION

"Among presumably normal adult subjects any $F\%$ between 50 and 80 invariably corresponds to signs of inflexibility, or, in clinical terms, constriction with compulsive elements" [7, p. 234]. Referring to Figure 1, the reader may note that the break in the figure significantly occurs at an $F\%$ of 50, thus providing experimental verification for Klopfer's hypothesis. In general the higher the $F\%$ for any given subject in the present sample, the greater his ability to cope with the situation of stress. But at the breaking point of 50 per cent the curve falls back upon itself, and the reverse of the above generalization becomes true: as the $F\%$ increases above 50, the subject becomes less able to cope with the situation of stress. Klopfer uses the descriptive terms "inflexibility" and "constriction" in describing such subjects, and the writers have interpreted these characteristics in terms of adaptation—specifically, adaptation to a testing situation involving stress. There is good evidence [1, 10] that the Information Test is particularly susceptible to emotional blocking, and, therefore, it might be expected that individuals who are susceptible to stress would give ready indication of it on the Information Test. The results in this direction reported here are really not so surprising.

[3] The present writers have striking, if not significant, evidence that the exceptionally young and old subjects in the sample have great tendencies to suffer breakdowns of intellectual control under stress. The age group between 25 and 40 years shows considerably better control in this respect. Such evidence served as further reason for restricting the age group in the final sample studied.

[4] See the discussion of the reliability of percentages and ratios by Cronbach [3].

But is it possible to generalize beyond this information and to state that *F%* might serve as a predictor of emotional blocking in virtually any behavioral situation where stress is involved? Obviously, such a broad statement cannot be verified by the present study, but it was for exactly the purpose of offering a partial answer in this direction that the digit-symbol test was also employed. The digit-symbol test is usually described as a measure of psychomotor speed, and the finding of the present paper that there was just as significant a deficit on this subtest as on the Information Test may be a first step in the direction of discovering that the predictive powers of *F%* are not limited to one kind of situation.

Klopfer does not consider the cases utilizing few *F* responses, but the present study indicates that a break at an *F%* of 30 may also be significant. Subjects falling below an *F%* of 30 indicate a trend in the direction of a greater maintenance of intellectual control under stress as fewer *F* responses are used. One possible explanation for this deviation lies in the utilization by the low *F* subjects of other Rorschach responses, probably in the direction of a greater use of color and movement, indicating a full and expansive personality rather than a constricted one. The real importance of the form-stress curve, however, is that with increasing use of *F%* there is an increasing ability to utilize intellectual factors in the face of emotional demands created by stress, and that the curve may thus give an estimate of just how susceptible the personality is to breaking down in stress-producing situations. Considering that the stress used in the present study was only of the mildest verbal type, it is even more remarkable that predictive powers of the curve are as great as they are.

One further word might be mentioned in regard to the age factor. The present paper indicates that as far as the meaning of *F%* is concerned, the older subjects cannot be handled in the same way as the younger subjects. Of course, most studies on normal subjects are based on a college population largely because of its availability, and this fact may be obscuring the importance of the age factor in Rorschach interpretation. In this connection it would have been of interest to correlate age, form, and the behavior criterion, but unfortunately it is not possible to fit *eta* into a multiple correlation formula.

SUMMARY AND CONCLUSIONS

Using 108 college subjects between the ages of 17 and 29 years, the relationship between certain Rorschach factors and stress was investigated. The Rorschach scores were obtained on the Harrower-Erickson Multiple Choice Group Rorschach using a modified method of scoring. A measurement of stress was obtained by the administration of a digit-symbol test and the Wechsler-Bellevue Information Test both before and after verbally induced stress. The digit-symbol test was used on 40 of the subjects, while the remaining 68 were scored on the Information Test. The differences between the two tests were converted into *T*-scores so that the data could be combined for the entire group.

The conclusions follow:

1. A curvilinear relationship was established between *F%* and the alterations in behavior caused by stress.

2. The hypothesis that a similar relationship would hold between $M/M + FM$ and changes in behavior under stress could not be verified in this study. The results, though not statistically significant, were in the expected direction.

3. People above 29 years of age made such restricted use of *F* that it is not possible to predict behavior under stress for the older age group from the *F* factor in the situation studied.

Received September 6, 1950.

REFERENCES

1. CRONBACH, L. J. *Essentials of psychological testing.* New York: Harper, 1949.
2. CRONBACH, L. J. "Pattern tabulation": A statistical method for analysis of limited patterns of scores, with particular reference to the Rorschach test. *Educ. psychol. Measmt,* 1949, **9,** 149-171.
3. CRONBACH, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.,* 1949, **46,** 393-427.
4. DERNER, G. F., ABORN, M., AND CANTER, A. H. The reliability of the Wechsler-Bellevue subtests and scales. *J. consult. Psychol.,* 1950, **14,** 172-179.

5. HARROWER-ERICKSON, MOLLY R., AND STEINER, MATILDA E. *Large scale Rorschach techniques.* Springfield, Ill.: Charles C Thomas, 1945.

6. HERTZ, MARGUERITE R. Rorschach: Twenty years after. *Psychol. Bull.*, 1942, **39**, 529-572.

7. KLOPFER, B. AND KELLEY, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.

8. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.

9. MONROE, RUTH L. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. *Appl. Psychol. Monogr.*, 1945, No. 7.

10. RAPAPORT, D. *Diagnostic psychological testing.* Chicago: Yearbook Publishers, 1945. Vol. I.

11. SARGENT, HELEN D. Projective methods: Their origin, theory, and application in personality research. *Psychol. Bull.*, 1945, **42**, 257-293.

12. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

13. WECHSLER, D. *The Wechsler-Bellevue Intelligence Scale. Form II. Manual for administering and scoring the test.* New York: Psychological Corp., 1946.

14. WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.

COLOR AS A DETERMINANT OF RESPONSIVENESS TO RORSCHACH CARDS IN SCHIZOPHRENIA[1]

SAMUEL L. BUKER
NORTHWESTERN UNIVERSITY
AND
MEYER WILLIAMS
VA HOSPITAL, DOWNEY, ILLINOIS

THERE are essentially two related hypotheses involved in relation to the use of color in the Rorschach test. The first hypothesis is that color gives rise to characteristic types of responsiveness, as may be noted in the differential scoring of direct color responses, the 8-9-10 per cent, and the indicators of color shock. The second hypothesis is that these types of responsiveness are related to specific personality variables as indicated by the relating of color to affect and emotional control. It is the first hypothesis which was the primary concern of this study.

While a few critical studies [9, 12] indicating a relationship between Rorschach color responsiveness and specific personality variables exist, critical studies which directly indicate that Rorschach color gives rise to characteristic responsiveness are absent. However, related investigations in this area have been made. For instance, Siipola [10] examined the relation of color to reaction time by presenting to her subjects specific portions of the chromatic cards in both chromatic and achromatic form, and found a significantly longer reaction time for the chromatic than for the achromatic. After examining the responses she decided that congruency of color with the concept suggested by the form of the area was the determining factor. Rabin and Sanderson [7] believe their study shows that color shock may be due to greater difficulty of cards, or greater potentiality of certain cards for responses. Wallen [11] found differences in responsiveness from colored card to colored card and thinks that color most affected response to Card II, but that shape most affected response to Card IX. Sappenfield and Buker [8] were unable to find a significant difference in *8-9-10%* when both a chromatic series and an achromatic series were presented to the same subjects in group form. The results of Lazarus [6], also using a group method of presentation of both an achromatic series and a chromatic series to the same subjects, suggest that the entire concept of color shock is questionable.

Taken as a group, the above studies suggest that there are variable relationships from colored card to colored card among the stimulus factors. Moreover, it is possible that the variable relationships may be a function of differences in form, hue, shading, or any combination of these. To be sure that interpretations of responsiveness based on the color in the Rorschach test are correct, it is necessary to know in what ways and to what extent color actually affects responsiveness, and in particular its effects on those scoring categories other than the direct color categories (*FC*, *CF*, and *C*).

It was the purpose of this study to make a direct attack on the problem by specific variation of the color variable while controlling other pertinent variables. Since much of the rationale of the value of color in the Rorschach

[1]Published with permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or conclusions drawn by the authors. Acknowledgement is due to Dr. W. A. Hunt and Dr. T. W. Richards of Northwestern University for direction and assistance in planning and completion of this experimental inquiry. The authors also wish to express their appreciation for the cooperation extended by Drs. B. S. Cane, Manager, A. Rodriguez, Chief of Professional Services, E. T. Sheehan, L. H. Rudy, R. I. Cutts, and J. R. Franco of the VA Hospital, Downey, Illinois.

test stems from what is presumed to be characteristic responsiveness of the various clinical groups, this experiment dealt with one of these groups, namely, the schizophrenic. In general, the schizophrenic is supposed to react with less inhibition to color on the Rorschach [2, p. 215; 5, pp. 356-357], and for this reason provides a favorable opportunity for evaluating the role of color. Thus, this experiment investigated the ways and the extent to which color in the Rorschach test is a determining factor in producing responsiveness from schizophrenic patients.

PROCEDURE

In general, procedure consisted of administering the regular series of Rorschach cards and a photographically reproduced achromatic series in counterbalanced order to a group of schizophrenic patients.

Materials. Materials consisted of the regular series of Rorschach cards (Series C) and a photographically rendered achromatic series (Series A), copying the regular series. There are four variables within the colored Rorschach stimulus cards. These are (1) form, (2) brightness, or shades of gray from black to white, (3) hue, or pure color, and (4) contrast resulting from hue-brightness combinations. In making a straight transition to photographic reproductions of the regular series of cards, brightness is held constant and hue is eliminated, but the contrast effect between hue and brightness changes when rendered as brightness only. Forms which are clearly delineated by light yellows and greens on the original series become almost indistinguishable from background on the achromatic reproduction. For this reason, special measures were taken to hold this contrast effect, and therefore form, relatively constant, although it meant slight changes in the brightness of various portions of the card. This was found to be necessary only for Cards VIII, IX, and X. The other cards were straight transitions.

Population. The experimental group was composed of 21 male white schizophrenic patients at The Veterans Administration Hospital, Downey, Illinois. Diagnosis was determined by complete agreement of a psychiatric staff (at least three psychiatrists and one psychologist) that each subject at the time of testing fell in the desired classification as described by the Veterans Administration Technical Bulletin 10A-78 [13]. To insure representativeness across the general schizophrenic classification, the subjects were subdivided as follows: six paranoids, six simples, six hebephrenics, and three catatonics. (Only three catatonics could be found who complied with all controls.) The mean age of the group was 29.4 years, and the age range was 20-40 years. The mean number of years the patients have carried their present diagnoses was 3.5, with a range of 0-8 years. No patients were used who were undergoing active therapy, receiving sedation, or who had had shock therapy less than six months previously. In addition, no patients with even suggestions of serious alcoholism or organic involvement in records and case histories were used. All subjects were tested for color blindness. Intellectual functioning of the subjects was measured by means of the CVS Abbreviated Intelligence Scale [3, 4], and only those patients were used whose converted scores, when prorated, were comparable to Wechsler-Bellevue Verbal Scale IQ's of 82-120. The mean estimated IQ of the group was 99.1 (SD 12.5). The mean number of years of education of the group was 11.1, with a range of 6-16 years.

Method. Half the members of each subgroup of six and two members of the subgroup of three catatonics, by random assignment, took Series C first, and the other subjects took Series A first. Of the total group of 21, 11 took Series C first. Each subject took the alternate form approximately one month after taking the first series. The mean number of days between tests was 33.1, range, 21-53 days.

All tests were administered, scored, and recorded by the senior author, using the scoring technique and the standard conditions of administration recommended by Beck [1]. In addition the records were scored for animal and inanimate movement (*FM* and *m*) as recommended by Klopfer [5]. Scoring was checked by the other investigator. Beck's instructions and a standard inquiry recording system were strictly followed for each series. Each subject was rated for cooperation, motivation, attention, and contact after each test. The two ratings for each subject were accomplished independently of each other.

RESULTS

In this section, the analysis of counterbalancing for control of test-retest effects is presented first. Following this is the analysis of the effects of removal of color, comparing Series C with Series A, regardless of position in sequence.

Test-Retest Effects. To examine the control for test-retest effects the distributions of scoring categories for all 20 cards were compared between the 11 subjects who took Series C first and the ten who took Series A first. A four-fold table and chi-square method of testing significance of differences was employed. This test was accomplished for 17 major scoring categories for each of three different sets of distributions: (1) for all 21 subjects, (2) for the 11 subjects with the highest response totals on both series, six of the subjects who took Series C first versus five who took Series A first (these individuals were constant throughout the comparisons for all scoring categories), and (3) for the ten subjects with the lowest response totals on both series, five each, for each series first. Cutting points for each of the three four-fold tables per scoring category were determined by taking that score providing whole numbers of cases which was closest to the median of the combined distributions. In case of tie the upper score was used.

For all of the above breakdowns, there were *no* chi-square values at or beyond the .05 level significance. The chi-square value for the sum of the *Y* scores for those patients with the lower response totals was the only one that approached significance ($.10 > P > .05$). An examination of the medians suggested considerably fewer *Y* responses for the group taking Series A first (median 3.5) than for the group taking Series C first (median 7.0). However, an examination of the cutting points suggested a definite relationship between this variable and total responses, the cutting point for the group with high response totals being 11.5 *Y* responses, while that for the group with low response totals was 3.5 *Y* responses. Even though chi-square comparing the distributions of *R* for the two sequences of *twenty* cards for these individuals did not approach significance ($\chi^2 = .417$), when comparisons of *R* were made between the four distributions of *ten* cards each, chi-square values of 3.75 ($.10 > P > .05$) were found for Series C first versus Series A first, and for Series A second versus Series A first. In each case the group taking Series C first had the higher response totals. This seems to be sufficient evidence for concluding that the difference in *Y* scores for the low response group can be partially explained on the basis of a difference in *R*. Granting this, it appears that the counterbalancing was an effective control of test-retest effects.

The Effects of Removal of Color. Statistical treatment of the data, checking for significance of differences and significance of relationships for the scoring categories for the two series was accomplished as follows. Since correlations between the two series were possible (each subject took each series), the data were examined in four-fold contingency tables. Two chi-squares were computed from each table: the first, a chi-square testing significance of change, and the second, a chi-square testing significance of association. From the association chi-square a coefficient of contingency was directly computed. The chi-square of change permits evaluation of changes in the distribution when the medians are different, and the chi-square of association and the coefficient of contingency permit an estimation of relationship. The combination of the two permits rather complete evaluation of differences or similarities of the two distributions. Medians and cutoff points were computed in the same way as for the treatment of test-retest effects. These tests were made for each scoring category for: (1) the entire distributions for each series, 21 cases in each, (2) the distributions of the 11 subjects with the highest *R* (hereinafter referred to as *UR*), and (3) the ten subjects with the lowest *R* (hereinafter referred to as *LR*).

The entire process, above, was completed for scoring category totals for (1) all ten cards, (2) the five cards which are achromatic on each series, and (3) the five cards which are achromatic on Series A and chromatic on Series C.

Table 1 presents the medians, chi-squares testing for significance of change (hereinafter referred to as χ^2_c), and coefficients of contingency (hereinafter referred to as *CC*) comparing Series C with Series A for the distribu-

TABLE 1

MEDIANS, CHI-SQUARES OF CHANGE, AND COEFFICIENTS OF CONTINGENCY COMPARING SERIES C WITH SERIES A FOR ALL TEN CARDS

Score	Medians			Chi-squares of Change			Coefficients of Contingency		
	Total Group	Series C	Series A	Total Group	Adj. for *R*		Total Group	Adj. for *R*	
					UR	*LR*		*UR*	*LR*
R†	15.8	18.0	15.0	‡			.47	.27	
W	6.9	6.7	7.3				.23	.27	
D	8.5	9.0	8.0				.53		.40
Dd	1.0	1.0	.7				‡		
S	1.4	1.4	1.3				.23	.41	
M	1.0	1.0	.7				.69		.52
M+	.7	.8	.6				.44		.52
FM	.9	1.0	.9		1.33		.39	.33	.34
Y	3.0	2.7	3.8		1.33		.25	.35	
V	.3	.3	.4						
F	10.1	9.8	10.3				.23		.37
F+	7.5	7.0	8.0				.40		
F+%	71.2	67.0	74.8	2.25			.50		
H	1.1	1.2	1.1				.43		.50
Hd	.9	.9	1.0				.39	.27	.37
A	6.3	6.0	6.4			1.33	.46	.27	.30
Ad	1.7	1.9	1.4				.25	.54	
A%	46.0	48.0	41.0				.40	.55	.37
An	1.1	.7	1.3				.30		.62
Bt	.4	.4	.4				.36		.57
Ge	.5	.8	.4	3.20*	1.33		.46	.27	.34
P	4.3	4.4	4.1				.53	.55	.37
Z	27.0	30.0	22.6	3.20*			.46	.27	
8-9-10%	30.8	33.3	30.4				.26		
T/1R"	17.0	19.3	14.0	4.17*	1.33	2.25	.41	.35	
T/R'	.8	.8	.8		1.33		.23	.35	

†"Number of" unless otherwise indicated.

‡Only χ^2_c's above 1.00 and only *CC*'s above .20 are listed.

*χ^2 of 3.84 = P at .05 level, of 6.64 = P at .01 level.

tions of various scoring categories for all ten cards for (1) all subjects, (2) the 11 subjects with the highest *R* (*UR*), and (3) the ten subjects with the lowest *R* (*LR*). For *most* of the scoring categories the table shows no χ^2_c which suggest significant changes. However, it is to be noted that three scoring categories have χ^2_c approaching significance (*Ge*, *Z*, and *T/1R"*).

The *Z* scores (organization scores) are most certainly related to *R* (cutting point for *UR* was 35.75, for *LR*, 20.75), and when the adjustment is made for this relationship the χ^2_c values drop below 1.00, suggesting that no significant change in *Z* occurred. The category *Ge* (geographical content) also is related to *R*, and when a correction for this factor is introduced the suggestion of a significant change disappears.

This leaves *T/1R"* (average time per first response) for the total distribution as the only possible significant change for all ten cards when color is removed ($.05 > P > .02$). The cutting scores (16.5 for *UR*, 18.0 for *LR*) suggest only a very slight relationship, if any, between *R* and *T/1R"*. The *CC* for the entire group (.41) suggests some degree of positive relationship between series, and that this index has some reliability. Inspection of the actual data shows that *all* changes across cutting points for the total group, *UR*, and *LR* occurred from above the median on Series C to below the median on Series A and that in each case approximately one-third of the cases was

involved. It is felt that in this case the null hypothesis may be rejected, and that removal of color significantly increased speed of reaction when all ten cards are considered.

Table 1 also shows *CC*, which range up to .69, suggesting varying degrees of relationship. (It should be mentioned that the highest possible *CC* for these four-fold tables ranges from approximately .69 for 21 cases to .63 for ten cases.) For example, such scores as *Dd* and *V* have *CC* below .20, suggesting little relationship for these variables between the two series, while such scores as *M* and *M*+ have *CC* of .69 and .44, respectively, for the total group and of .52 for *LR* but below .20 for *UR*. This suggests that a fair degree of relationship exists for these scoring categories between series, remaining relatively constant regardless of the total number of responses.

It will be noted that the median number of responses for Series C was 18.0 and for Series A, 15.0. It also happens that the median number of direct color responses (*FC*, *CF*, and *C*) on Series C was 2.9 (range 0-11). One might suspect that the dropping out of color responses accounts for the difference in *R* medians. As a check, the change in *R* from one series to the other was compared with the number of color responses present on Series C for each subject, and the resulting coefficient of contingency was —.0025, highly suggesting no relationship between these two changes.

The data for the five cards which are achromatic on both series received identical statistical treatment as did the data for all ten cards. Only two scoring categories displayed χ^2_c which suggested significance: *Ge* again, and the *F*+%. As was stated before, *Ge* is related to *R*, and when this is taken into account, the χ^2_c values drop from 4.17 ($.05 > P > .02$) for the total group to 3.20 ($.10 > P > .05$) for *UR*, and to below 1.00 for *LR*.

The χ^2_c for *F*+% for the total group and for *UR* (4.00 and 4.17) need further consideration. The medians for Series C and Series A in all card groupings revealed that in each case the median *F*+% for Series A was higher. It would be difficult to claim the change was due to removal of color, since the significant χ^2_c appear only when comparing the five cards which are achromatic on both series, and do not appear when comparing the five cards which are achromatic on one series and chromatic on the other. However, a possible explanation will be deferred to the discussion section.

It was also noted that *T*/*1R*″, suggested as being significantly affected by removal of color when comparing all ten cards, showed no chi-square values approaching significance when comparisons were made for these five cards which are achromatic on each series. This suggests that the increase in intial reaction time is a function of the chromatic cards.

The *CC* for the five cards which are achromatic on both series once again suggested markedly varying degrees of association between the two series for the different scoring categories. *D*, *M*, *A*%, and *P* displayed *CC* suggesting relatively constant association regardless of *R*, while *W*, *V*, *F*+ and *T*/*1R*″ all had *CC* below .20, regardless of *R*.

The data for the five cards which are chromatic on one series and achromatic on the other received the same statistical treatment accorded to the above card groupings. The most important fact revealed in examination of these data was that only one χ^2_c, that for *T*/*1R*″ for the total group, approached a value which suggested significance (3.13, $.10 > P > .05$). This is consistent with the above suggestion that the differences in *T*/*1R*″ were a function of the chromatic cards.

The *CC*'s, although suggesting a greater variation in the degrees of association between the two series for the different scoring categories, nevertheless showed findings relatively consistent with those previously noted. This greater variation would be expected, since if whatever reliability exists for the regular series is affected by removal of color, the effect should be most noticeable in this card grouping.

Qualitative Findings. A comparison by inspection of the two protocols for each subject revealed only one relatively consistent trend which appeared related to color. Ten of the 21 subjects appeared to have exchanged *C* responses on Series C for *Y* responses on Series A. This did not appear to be related to the series taken first. The majority of those subjects made several *Y* responses to Series C

and made approximately as many more *Y* responses on Series A as they had *C* responses on Series C. It was noted that the median number of *Y* responses for all ten cards was 2.7 for Series C and 3.8 for Series A. This gain was most noticeable on the cards which are chromatic on Series C, where the medians were 0.5 for Series C and 1.2 for Series A.

DISCUSSION

The results of this study suggest that for the five chromatic cards the *only* significant effect of removal of color was a decrease in reaction time. This change in reaction time was also significant when all ten cards were considered as a unit, but not when the five achromatic cards were considered as a unit. This suggests that the presence of color has a depressant effect on reaction time, which finding concurs with that of Siipola [10].

The only other possible significant effect of removal of color suggested is that the *F*+% for the five achromatic cards is somehow influenced, without the *F*+% for the five chromatic cards becoming involved. Since a lack of reliability of this score is suggested, it would be easy to offer an explanation in terms of the misfortunes of sampling and to accept the null hypothesis. However, another possible explanation must be mentioned, that color indirectly exercises a delayed depressant effect on the *F*+% for these cards. This receives some support from the fact that the median *F*+% for Series A for each card grouping is higher than that for Series C. It is suggested that the focal point of this effect may be the change from the partially colored Card III to the heavy gray on Card IV. However, due to the limited sample, no conclusive investigation of this possibility was made.

Probably the most important finding suggested by the results is that for most of the scoring categories removal of color had no effect when considering all ten cards or the two groups of cards divided on the basis of the presence of color in the regular series. While it is possible that the limited number of subjects was not sufficient to reveal all effects of removal of color, at the same time, the lack of change, to a significant extent, is suggested in such scores as *P*, *M*, *M*+, and *A*% and to a lesser extent in many of the other scores. It is noted in particular that no significant change is suggested in the *8-9-10*% and in the usual indicators of color shock (other than *T*/*1R*" and possibly the *F*+%). These results might seem to concur with the thinking of Beck [2, p. 215] and Klopfer [4, p. 356 ff.], that the schizophrenic tends to respond less to color, i.e., is less responsive to the environment, than the normal. However, since these subjects did make direct responses to color (median *C* responses 2.9) and since these responses suggest poor control of affective response to exciting stimuli ($CF + C > FC$), there is reason to believe that these particular subjects should show other signs of disturbance of functioning when faced with the chromatic cards.

Two possible but not mutually exclusive explanations for the lack of change are suggested. The first is that the relationship between color and color shock cannot be completely evaluated on the basis of the card groupings used in this study. The second is that the role of color is not as important as is generally believed. It is very likely that color is primarily important only as part of the total gestalt of the card, as other studies have suggested [6, 7, 10, 11], and that the disturbance of functioning is due to something other than simply the presence of color. It is also possible that the sequence of the cards is even more important than is generally believed. Finally, it must be mentioned that even the concept of "shock" may be the result of little more than a measure of the differences in difficulty of the separate cards.

CONCLUSIONS

From the results and discussion of this study it is felt that for an adult male schizophrenic population, the following conclusions about the ways and extent that color influences responsiveness to the Rorschach test are warranted:

1. In general, color, as a discrete variable, does not materially influence responsiveness to the Rorschach.

2. The presence of color increases initial

reaction time to the five chromatic cards, but not to the five achromatic cards.

3. There is some possibility that color may also exercise a depressant effect on the $F+\%$ for the five achromatic cards, but not affect the $F+\%$ for the five chromatic cards. If this is so, it is felt that the focal point for this change may be the transition from Card III to Card IV.

Received September 29, 1950.

References

1. Beck, S. J. *Rorschach's test. I. Basic processes.* New York: Grune & Stratton, 1944.
2. Beck, S. J. *Rorschach's test. II. A variety of personality pictures.* New York: Grune & Stratton, 1945.
3. Hunt, W. A., and French, Elizabeth G. A second fifteen word vocabulary test. *J. consult. Psychol.*, 1949, **13**, 124-126.
4. Hunt, W. A., French, Elizabeth G., Klebanoff, S. G., Mensch, I. N., and Williams, M. Further standardization of the CVS Individual Intelligence Scale. *J. consult. Psychol.*, 1948, **12**, 355-359.
5. Klopfer, B., and Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
6. Lazarus, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, **44**, 506-516.
7. Rabin, A. I., and Sanderson, M. H. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.*, 1947, **3**, 216-225.
8. Sappenfield, B. R., and Buker, S. L. Validity of the Rorschach 8-9-10 per cent as an indicator of responsiveness to color. *J. consult. Psychol.*, 1949, **13**, 268-271.
9. Sarason, S. B., and Potter, E. H. Color in the Rorschach and Kohs Block Designs. *J. consult. Psychol.*, 1947, **11**, 202-206.
10. Siipola, Elsa M. The influence of color on reactions to ink blots. *J. Personality*, 1950, **18**, 356-382.
11. Wallen, R. The nature of color shock. *J. abnorm. soc. Psychol.*, 1948, **43**, 346-356.
12. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.
13. *Vet. Adm. tech. Bull.* 10 A-78, October 1, 1947.

THE AMBIGUITY VALUES OF TAT CARDS[1]

SIDNEY W. BIJOU AND DOUGLAS T. KENNY
UNIVERSITY OF WASHINGTON

PROJECTIVE techniques such as the Thematic Apperception Test are based upon several assumptions. One is that in the construction of stories around ambiguous picture stimuli, usually referred to as fantasy production in a test situation, the individual organizes his material from his own personal experiences as well as from his present feelings, sentiments, needs, and strivings. Another assumption, which incorporates the first, is that the more ambiguous the picture stimuli, the greater the apperceptive distortion, i.e., the more deviant the interpretation of perceptions, the more revealing of the personality structure. A generalized statement of this relationship is given by Murphy [8, p. 351]: "The more tightly organized, the better structured the world pattern, the less it can be controlled by the inner pattern; the more lowly structured, the more the inner pattern controls it."

Murray [11] applied this principle in constructing the Thematic Apperception Test. He selected pictures so that the first ten of the series would be fairly well structured while those in the second ten would be "more unusual, dramatic and bizarre than those of the first." This difference, together with a modification in the test instructions to encourage freer association, was designed to have the second series of cards produce more fantasies yielding "richer" personality material.[2]

The assumption of an inverse relationship between stimulus structure and extent of the personality factors "injected" into the response is also currently held by many of the investigators interested in the influence of need on perception. For example, Bruner [2, p. 161], in discussing how the perceptual field comes to be organized, writes:

> Generally speaking, the less the structuredness or univocality of the stimulus input, the more striking the role of directive, nonsensory factors will be in determining perceptual organization. An ink blot or a cloud picture is a nice instance of the kind of equivocal stimulus which maximizes the role of personal or directive factors in perceiving.

Recently, Dollard and Miller [3] employed this hypothetical relationship together with the assumption of perceptual generalization to account for hallucinatory behavior. They maintain that with heightened motivation the individual will generalize his perceptual responses to stimuli which are either ambiguous or are similar to the original stimulus.

Although the above relationships between stimulus ambiguity and apperceptual and perceptual behavior have been widely accepted, an evaluation of experiments in this area leads one to believe that further evidence should be welcomed. At this particular time, it seems that considerable information is needed to work out statements on the relationship between stimulus ambiguity on one hand, and perception, apperception, and need states on the other. Our concern is with stimulus ambiguity and apperceptual phenomena.

Weisskopf [13] performed one of the few studies on the relationship of the stimulus and the apperceptive response when she investigated the degree of fantasy elicited by TAT pictures as a function of stimulus brightness and ambiguity. The amount of fantasy elicited

[1]This investigation was supported in part by funds from the University of Washington Research Committee.

[2]Although it is possible that more meaningful material from the point of view of personality analysis is elicited from the second series of pictures, one is not sure to what extent this is due to the more dramatic qualities of the pictures in the second series or to the change in the instructions.

was measured by "transcendence indices," which were defined in terms of the mean number of responses "which go beyond pure description." There were three experiments. In the first, stimulus ambiguity was varied in terms of intensity or brightness. The results indicated there was no significant difference between amount of fantasy elicited by pictures presented under conditions of reduced and of normal brightness. In the second experiment, stimulus ambiguity was varied by presenting complete and incomplete tracings of a set of pictures. She found that the more structured pictures, those with complete tracings, evoked reliably more fantasy material. In the third experiment, stimulus ambiguity was manipulated in terms of exposure time, one set being exposed for 0.2 second, the other for 5 seconds. The pictures exposed for the longer period elicited reliably higher fantasy scores. Evidence from this study, therefore, seems to indicate that stimulus ambiguity, defined in terms of physical properties, does not appear to be inversely related to amount of fantasy production. These findings may be used to evaluate the proposition under consideration provided one assumes that (*a*) stimulus ambiguity is properly defined in terms of physical properties, and (*b*) there is a direct relationship between amount of fantasy and amount of personality factors introduced. Both these assumptions may be questioned.

Considerable research is necessary to accumulate evidence that will determine the validity of this assumption which seems to be basic to the interpretation of the pictorial type of projective material. Data of this nature should contribute to a personality theory as mentioned above and should also have practical significance. Clinical psychologists, for example, should know what types of pictures yield the greatest amount of personality material. Such information would assist in diagnostic procedures and would, in the long run, aid the further development of projective techniques. If it were clearly established that more ambiguous stimulus material yields more significant information, then much trial and error could be eliminated in the construction of projective techniques.

THE PROBLEM

The present study is designed to investigate further the assumption of a direct relationship between the ambiguity qualities of picture stimuli and the extent of personality factors revealed in the elicited fantasy. In planning an approach to the problem, such questions as these were encountered: How is stimulus ambiguity to be measured? What measures can be used to indicate the amount of "personality structure" revealed or extent of "inner pattern in control"? What is the most efficient procedure to relate stimulus and personality measures in such a problem?

After considering these questions, we decided to proceed with an experimental design involving: (*a*) three sets of TAT pictures varying in degree of ambiguity, (*b*) three comparably trained test administrators, (*c*) three groups of subjects from a common population, and (*d*) fantasy behavior scales which would permit applying statistical tests (analysis of variance) to the relationships between such scale values and stimulus ambiguity measures. The experiment was therefore conducted in three parts: (*a*) scaling TAT pictures on the basis of an ambiguity dimension, (*b*) developing rating scales for quantifying TAT protocols, and (*c*) analyzing the relationships between stimulus ambiguity and fantasy scale measures.

Scaling TAT cards or any other set of pictures devised to instigate fantasy productions creates a problem for study when it is believed, as we do, that the ambiguity dimension should be defined in terms of responses rather than some physical property of the pictures. The present paper is confined to an exposition of the problem of establishing ambiguity values for TAT pictures. Subsequent reports will describe the construction, reliability and use of the fantasy scales, and the finding on the relationship between stimulus ambiguity and variations in fantasy scale value.

METHOD

Material. The materials consist of the 21 TAT cards used with male clients, these having been selected because only male subjects were to be employed in the subsequent study.

The designations of these cards are: 1, 2, 3BM, 4, 5, 6BM, 7BM, 8BM, 9BM, 10, 11, 12BG, 12M, 13MF, 13B, 14, 15, 17BM, 18BM, 19, and 20. The blank card (number 16) was omitted because it appeared to be on a different stimulus continuum from that of the other cards. On all cards, except 16, the individual is allowed to organize the external cues and so guide his projections, but on Card 16 his responses are in no apparent way "externally" determined, but almost entirely, if not completely, determined by "internal" factors. Thus, to the extent that on this card no impoverished stimulus situation or reduced cues are provided to contribute to the determination of the perceptions, whatever projections occur must result from what Bruner calls "behavioral determinants," such as needs, motives, and values. It was because of this consideration that Card 16 was excluded.

Subjects. To judge the cards we used 51 college students from one of the introductory psychology courses given during the summer of 1949 at the University of Washington. Of this number, 28 were males and 23 females. The age and educational levels of the subjects are summarized in Table 1.

TABLE 1

AGE AND EDUCATIONAL LEVEL OF THE SUBJECTS

Group	*N*	Mean Age	Mean Ed. Level (Years in Univ.)
Males	28	27.7	3.2
Females	23	25.5	2.6

Procedure. Each subject (judge) was invited into an office and asked to participate in an experiment in which he would be asked to make judgments about pictures. He was then given the following instructions to read carefully:

General Instructions

You have here a number of samples of pictures from which a psychological scale is to be constructed. For the purposes of the scale you will be asked to compare the different pictures and judge the *ambiguity* of them. You are to regard them as social pictures illustrated in a story book.

By *ambiguity* we mean *that which is open to a number of possible interpretations.* Base your judgment on this factor only. Be careful to eliminate such factors as the particular associations, pleasant or unpleasant, which you may have with certain pictures.

Directions for Ranking Pictures

The pictures are spread out in front of you in a random fashion. You are to arrange them in sequence or rank order according to their degree of ambiguity.

Rank as number one the picture that is *least open to a number of possible interpretations* and place it to the extreme left; next find the second least ambiguous picture and place it beside the first; then the third least ambiguous picture and place it by the second and so on down to the picture that is *most open to a number of possible interpretations,* which will appear at the extreme right.

After you have laid out the whole sequence before you, you may wish to change the order of the pictures. You may change the order of the pictures as many times as you like to obtain your final ranking.

After the subject finished reading the instructions, he was asked whether further clarification was necessary. In response to questions, the experimenter rephrased the pertinent part of the instructions or referred directly to the instruction sheet. When the experimenter was certain the subject understood his task, he told him he could take as much time as he needed to rank the pictures and then left the room.

TABLE 2

MEANS, STANDARD DEVIATIONS, AND DIFFERENCES BETWEEN MALE AND FEMALE JUDGMENTS OF AMBIGUITY ON 21 TAT CARDS

Picture Designation	Males ($N=28$) Mean	SD	Females ($N=23$) Mean	SD	Mean Diff.	*t*
1	7.53	5.22	6.26	3.62	1.27	0.96
2	8.35	5.77	8.56	4.67	0.21	0.13
3BM	12.28	4.65	11.17	4.79	1.11	1.35
4	11.28	4.68	12.95	4.64	1.67	1.24
5	14.00	5.02	12.13	5.47	1.87	1.50
6BM	12.42	4.90	12.52	4.09	0.10	0.07
7BM	13.28	4.66	11.47	4.61	1.81	1.36
8BM	11.03	5.24	12.04	4.39	1.01	0.72
9BM	7.50	4.96	8.60	5.40	1.10	0.74
10	12.03	4.73	4.82	5.06	2.57	1.83
11	13.89	7.05	13.08	7.98	0.81	0.37
12BG	5.82	6.96	4.17	4.94	1.65	0.93
12M	11.00	5.90	12.87	4.08	1.87	1.26
13MF	11.72	5.15	12.04	5.22	0.32	0.21
13B	6.71	4.90	9.04	6.11	2.33	1.48
14	9.46	5.40	8.61	5.57	0.85	0.54
15	11.57	5.83	13.87	6.93	2.30	1.26
17BM	9.00	5.46	8.57	5.54	0.43	0.27
18BM	15.25	4.69	14.35	5.51	0.90	0.61
19	14.64	7.00	15.04	6.99	0.40	2.01
20	12.25	4.91	13.65	5.03	1.40	0.98

RESULTS

Sex Differences. There were two reasons why we felt it necessary to examine our data for the possibility of sex differences in the ranking of the pictures. First, as our later study was to employ only males, sex should not affect the ambiguity values of our cards. To treat male and female rankings as being alike assumes that the concept of ambiguity is the same for both sexes. Second, it was essential to analyze possible sex differences because only male cards were used in the series judged and there was a possibility that females might rank the pictures differently from males.

As is shown in Table 2, the differences between the sexes were not statistically significant. We may therefore conclude that there is a consistent tendency for the males and females in our group to judge the pictures in a similar fashion.

Although statistical comparison of differences between the sexes for each card has shown agreement in their judgments, it is of interest to compare the rank orders assigned to each picture by the sexes. When the rankings of the pictures for the males and females are correlated a ρ of .75 is obtained. Following Peatman's suggestion [12] for treating ρ as a product-moment correlation coefficient and then using Fisher's z transformation function, we find this ρ is significantly different from zero at the 1 per cent level. It is to be noted that the ranks assigned to the cards by the males are in high agreement with the ranks assigned by the females. This male and female agreement may be taken to mean that there is a "psychodynamic" neutrality between the sexes on judging such ambiguous stimulus cards as the TAT.

TABLE 3
RANK ORDER, MEDIANS, MEANS, AND STANDARD DEVIATIONS OF 21 TAT CARDS JUDGED AS TO DEGREE OF AMBIGUITY

Rank	Picture Number	Median	Mean	Standard Deviation
1	12BG	1.00	5.07	6.19
2	1	6.00	6.96	4.80
3	2	7.00	8.45	5.31
4	9BM	7.00	8.00	5.19
5	17BM	7.00	8.80	7.30
6	13B	7.00	7.76	5.60
7	14	8.00	9.08	5.49
8	10	10.00	11.23	5.04
9	4	11.00	12.03	4.74
10	7BM	12.00	12.47	4.72
11	8BM	12.00	11.49	2.53
12	12M	12.00	11.84	5.24
13	13MF	12.00	11.86	5.19
14	20	13.00	9.07	5.49
15	3BM	13.00	11.78	4.74
16	6BM	14.00	12.47	4.55
17	5	14.00	13.15	5.31
18	15	15.00	12.60	6.45
19	18BM	17.00	14.84	5.10
20	11	18.00	13.52	7.63
21	19	19.00	14.82	7.00

Rank Order of 21 TAT Cards. The final mean and median ambiguity value, based on our 51 subjects, and the standard deviation for each card were then computed. These statistics, together with the rank order value of each card, based upon the median value, are shown in Table 3.

Judge Reliability. The method employed for estimating the reliability of the judgments is similar to determining split-half reliability. The subjects were numbered from 1 to 51, the odd-numbered subjects comprising one group and the even-numbered subjects the other group. After the rank order values assigned to each of the cards by both groups had been determined, a ρ between these ranks was computed. This ρ was .80, which is significant at the 1 per cent level. From this finding it appears that these students following the instructions given can make fairly reliable judgments on the ambiguity quality of TAT cards.

DISCUSSION

In testing the assumption of a positive relationship between stimulus ambiguity and extent of personality structure revealed in fantasy production, the question of whether stimulus abiguity is to be defined in terms of physical properties or in terms of behavior is of crucial importance. Some investigators such as Weisskopf [13] use physical dimensions to define the concept; others such as Bellak [1] imply that the definition must be in terms of group judgments; still others, including Murray [10], use both dimensions in their definition. Our approach, like Bellak's, was psychologically rather than physically oriented, our primary purpose being to develop a frame of reference for classifying the TAT pictures as to their degree of psychological ambiguity rather than to establish a system for identifying stimulus patterns on a continuum of physical "impoverishment." It followed that the man-

ner of determining the vagueness value of a constellation of stimuli evoking apperceptive reactions would have to be in terms of the consensus of a group from the population upon which we wish to test the assumption in question. Variation of some physical property of a card would not necessarily mean that it would be perceived as vaguer or clearer in its import. Physical ambiguity may not be the same as psychological ambiguity; in fact, there seems to be no reason to expect an isomorphic relationship between the two. In this sense, the present point of view is somewhat similar to the well-known distinction that Koffka [5] has made between the "behavioral environment" and the "real or physical environment." This position does not mean, however, that the physical characteristics of the picture are unimportant. On the contrary, we believe that considerable research is needed on the covariations between the physical and psychological aspects of sets of pictures used for projective analyses if we are to understand the projective process.

After deciding that stimulus ambiguity should be defined in terms of judgments, then the question arises as to how the concept should be defined for the judges, i.e., what should the specific instructions be. Unfortunately, one cannot turn to the literature for a standard meaning for this or related terms. As Luchins [7] recently pointed out, there is much need for clarification of such terms as ambiguous or unambiguous, well-structured or poorly-structured, and unstructured.[3] In the light of this situation, we decided to make an exploratory study of definitions. Our first attempt to describe the meaning of stimulus ambiguity was in terms of "the clearness of meaning" of

[3]For the purpose of this study, we employ the terms "vagueness" and "definiteness" to describe the psychological stimulus ambiguity continuum, vagueness and low structurization being used synonymously, and definiteness with high structurization.

TABLE 4
AMBIGUITY RANK ORDER VALUES FOR FIFTEEN CARDS SELECTED FOR STUDY

Final Rank	Picture Designation	Description
1	12BG	A rowboat on the banks of a woodland stream.
2	1	A boy and a violin which rests on a table in front of him.
3	2	Country scene: Young woman in foreground with books; in the background, a man working in the fields and an older woman looking on.
4	9BM	Four men in overalls are lying on the grass.
5	17BM	A naked man is clinging to a rope.
8	4	A woman is clutching the shoulders of a man whose face and body are averted as if he were trying to pull away from her.
9	7BM	A gray-haired man is looking at a younger man who is staring into space.
10	8BM	An adolescent boy with a barrel of a rifle at one side, and in the background is a dim scene of a surgical operation.
11	12M	A young man is lying on a couch with his eyes closed. Leaning over him is an elderly man, his hands stretched out above the face of the reclining figure.
12	13MF	A young man is standing with his head buried in his arm. Behind him is a figure of a woman lying in bed.
17	5	A woman is standing on a threshold of a half-opened door and looking into a room.
18	15	A gaunt man with clenched hands is standing among gravestones.
19	18BM	A man is clutched from behind by three hands.
20	11	A road between high cliffs, and obscure figures in the distance. On one side is the long head and neck of a dragon.
21	19	A weird picture of cloud formations overhanging a snow-covered cabin in the country.

pictorial representation. It seemed to us that a social situation or a pictorial representation of it could be judged high or low in ambiguity or structure depending on how clear the meaning of the picture appeared to a group of judges. The feasibility of the definition was evaluated in a pilot study involving ten college students. Considering the numerous questions asked after receiving the instructions, and the high variability for the mean rank values, it was apparent that this definition was not readily understood and was difficult to maintain as a criterion. In a second exploratory study, ten different judges were told to rank the pictures in terms of the estimated number of interpretations they thought would be possible for each picture. The results indicated that such a definition would be workable, and it was therefore adopted for use in the experiment described here.

It will be noted that total median rank values have been used to express the findings. We considered the median a better statistic than the mean for the final selection of our pictures because the equality of our unit of measurement is uncertain. Guilford [4, p. 78] has stated that the median is the most statisfactory average under such conditions.

> . . .We are not sure that all the units of our measuring scale are equal. This is particularly true in the psychological scaling methods of rank order and of equal-appearing intervals. In the former case, a number of judges have placed several objects . . . in rank order for some quality. Though the ranks are numerically equidistant, the things ranked probably are not. When combining ranks for any one object, we do less violence to the measurement if we find a median rather than a mean.

Cards 12BG, 1, 2, 9BM, and 17BM had the lowest median rank values and were therefore considered to be least ambiguous or most highly structured. Cards 4, 7BM, 12BM, and 13MF were given intermediate rank values; and Cards 5, 15, 18BM, 11, and 19 had the highest median rank orders and therefore were judged to be of most ambiguous or low structurization value. Table 4 shows these findings in tabular form and also gives a brief description of each of the 15 pictures.

It will be of interest to compare the ranking found here with Murray's classification. Murray, you will recall, divided the cards into two sets of ten, cards numbered 10 and below being more structured than those numbered 11 and above. Placement was based on the physical characteristics and content of the card as well as on its diagnostic value. Comparisons between the two sets of rankings may be summarized as follows: (*a*) In our low ambiguity group, three of the cards are from Murray's first ten; (*b*) in the intermediate group, three are from Murray's first ten and two are from his second ten; and (*c*) in the high ambiguity group, four of the five cards are in Murray's second ten. Considering the differences in criteria employed by Murray and those used in this approach, the disparity is not too great.

Another comparison was made by a statistical analysis of Murray's placement of the 21 cards by computing the significance of the difference between the means of the first 10 cards and the last 11 cards. The null hypothesis to be tested was: The mean ambiguity value computed for the first 10 cards is not significantly different from the mean ambiguity value of the last 11 cards. These findings are summarized in Table 5. This t is significant at the five per cent level.

TABLE 5
SIGNIFICANCE OF DIFFERENCE BETWEEN MURRAY'S FIRST TEN AND LAST ELEVEN TAT CARDS

	Mean	SD	Diff.	t
First 10 cards	10.78	2.05	0.38	2.27
Last 11 cards	11.16	2.99		

Adopting a stringent level of significance, a P of .01, we may reject Murray's contention that the first 10 cards differ significantly from the last 11 cards in terms of ambiguity as we have defined it.

On the basis of our findings, the 15 cards described in Table 4 were selected for use in a subsequent experiment testing the relationship between stimulus ambiguity and extent of personality factors revealed.

SUMMARY

This study is concerned with establishing ambiguity values for TAT pictures as a first step in investigating the relationship between ambiguity of picture stimuli and the extent of personality structure revealed in the fantasy produced. Ambiguity is defined in terms of psychological rather than physical properties.

Fifty-one judges ranked 21 TAT cards (the general and male series) on the basis of their ambiguity, which was defined in terms of the estimated number of possible interpretations a picture evoked. A picture stimulating many interpretations was judged as vague or low in structure, while one eliciting one or few interpretations was judged to be definite or high in structure. The final ambiguity value of a card was expressed in terms of its median rank order value.

Statistical analyses of results revealed: (*a*) The multi-judge reliability of the ranking seems to be satisfactory ($\rho = .80$). (*b*) The rank order differences between the male and female judges were not reliably different.

The present findings on ambiguity were compared with Murray's division of the cards into two series of ten each. Judged by our definition of ambiguity, Murray's belief that cards numbered 10 and below are more structured than those numbered 11 and above is called into doubt.

On the basis of our present results, 15 cards varying in ambiguity value were selected for a subsequent study on the relationship between stimulus ambiguity and extent of personality revealed in the fantasy produced.

Received January 3, 1951.
Early publication.

REFERENCES

1. BELLAK, L. On the problems of the concept of projection. In L. E. Abt and L. Bellak, *Projective psychology*. New York: Knopf, 1950.
2. BRUNER, J. S. IV. Perceptual theory and the Rorschach test. *J. Personality*, 1948, **17**, 157-168.
3. DOLLARD, J., AND MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1950.
5. KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
6. KRECH, D., AND CRUTCHFIELD, R. S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
7. LUCHINS, A. S. The stimulus field in social psychology. *Psychol. Rev.*, 1950, **57**, 27-30.
8. MURPHY, G. *Personality*. New York: Harper, 1947.
9. MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310-339.
10. MURRAY, H. A. *Explorations in personality*. New York: Oxford Univ. Press, 1938.
11. MURRAY, H. A. *Manual for the Thematic Apperception Test*. Cambridge: Harvard Univ. Printing Office, 1943.
12. PEATMAN, J. G. *Descriptive and sampling statistics*. New York: Harper, 1947.
13. WEISSKOPF, EDITH A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.*, 1950, **29**, 407-416.

FACTORS AFFECTING THE CHOICE OF CERTAIN SZONDI TEST PICTURES

RICHARD WALLEN

CLEVELAND COLLEGE, WESTERN RESERVE UNIVERSITY

WHEN a person takes the Szondi Test, he is asked to choose pictures that he likes and pictures that he doesn't like. No other projective test uses this method. If this test works (and many people think it does), we ought to try to find out why it works. Why does preference for pictures show something about personality? The specialists who invented and developed the test have tried to answer this question, but their answer is not satisfactory in some ways.

The test pictures are photographs of patients in mental hospitals. These patients have been classified into eight different diagnostic groups, and there are six different patients pictured in each group. These 48 pictures are divided into six sets of eight pictures so that each set has one picture from each diagnostic category. A subject is given one set at a time and is told to choose the two pictures he likes best and the two he likes least. When he has finished choosing from all six sets, the examiner has a record of 12 preferred pictures and 12 rejected pictures. Then the examiner puts down the total number of choices and rejections for each of the eight diagnostic groups (called "factors" by Szondi). Interpretation is based on this pattern of scores. For dependable results, the test should be repeated from six to ten times. Then one can see which factors yield fairly constant scores and which ones yield changing scores.

Current ideas about why the test works may be divided into two groups. One group relates to the pictures; the other has to do with the psychological processes that lead people to choose or reject pictures. These theories raise questions about the psychology of preference that are interesting even aside from the value of the test itself.

Szondi specialists believe that the six pictures belonging to a given factor—for example, catatonic schizophrenics—have common characteristics that are not found in pictures belonging to the other factors. That is the reason for tabulating a subject's choices according to the diagnosis of the pictures he chooses. If you look carefully at the pictures, you may have some doubt about this principle. Photographs in a single factor show a good deal of difference in lighting and pose. The patients who are given a single label do not show the same facial expressions; and the amount and kind of clothing vary from picture to picture. Even though you agree that *something* is common to pictures within the same factor, you will probably wonder whether it really affects people when they make their choices. Too many other cues—and pretty obvious cues—are present. They could cover up any effect that the common feature might have.

Sometimes the same cues are found in most of the pictures in a single factor but are not found in the other factors. A good example is the fact that five of the six pictures of sadists are lighted by light sources below eye level. None of the other pictures was taken with this kind of lighting, as can be seen by inspection. Low lighting makes for queer, dramatic shadows. If this lighting affects people's choices, it could produce a homogeneous impression in these pictures that has nothing to do with sadism.

Deri has offered some interesting views on the second problem: why people choose or reject certain pictures. She says:

> We assume that the eight factors (diagnostic categories) in the test correspond to eight different need-systems in the organism. . . .The eight types of mental and emotional diseases represented in the test have to be thought of as expressing certain psychological needs in extreme form, which to some degree exist in everybody. . . . Depending on the degree of (or intensity of) the state of tension in each of the eight need-systems, the pictures representing the corresponding needs will assume valence character in various proportions. In this case the subject chooses

pictures from the factor corresponding to his own need in tension [1, p. 26].

Later she suggests that choosing a picture as liked shows "a conscious or unconscious identification with the motivational processes as depicted by the photographs of the respective factor," while disliking a picture indicates a counter-identification with these processes [1, p. 34]. What are these "motivational processes"? From what Deri says about the various factors, they must be processes that are found in patients with specific sorts of disorders. But how are these processes shown in the faces of the patients? We think it would be useful to find out what people think they see in these faces and how these characteristics are related to the relative popularity of the pictures.

Our purpose in this study was primarily to test the assumption that the catatonic and paranoid pictures produce different impressions on subjects. If grouping by diagnosis is defensible, it should easily be possible to show these two groups produce different kinds of judgments. For purposes of comparison we also wish to investigate differences in impression produced by a grouping based on definable stimulus characteristics of the pictures. Finally we shall present data showing the relationship of certain descriptive traits assigned to the pictures and subjects' preferences for them.

We chose to work with ego-vector pictures because scores based on them are of the highest importance in interpreting Szondi profiles. Deri devotes a little over a third of her book to a discussion of this vector. These 12 pictures are all of schizophrenic patients. Six of the patients were catatonic (these are called *k* pictures), and six were paranoid (the *p* pictures).

PROCEDURE

First, we asked 40 people to decide what trait-names applied to each of the 12 pictures. We furnished a list of trait-names for them to use. Their judgments enable us to describe each picture. The descriptions can be compared with those we would expect on the basis of present rationale.[1]

[1]The writer is grateful to Mr. George Lehner for helping to collect the judgments of the pictures.

Then we administered the Szondi Test to a completely new group to determine the popularity of the 12 pictures. The administration was individual and according to the standard method. For this study we used only the results from the initial testing. All our subjects were unfamiliar with the test and had never taken it before.

Judgments of the Pictures. Forty judges, each working alone, rated the photographs. Each judge was given a rating sheet with these instructions:

Here is a list of phrases describing personality characteristics. They have been arranged in pairs to make their meaning clearer. Look at each photograph you are shown and decide what kind of a person is pictured. Check all the phrases which apply to the picture. You do not have to check one of each pair, because some pairs will not apply to the picture. Just check those which seem to fit best.

Below these instructions the pairs of phrases were printed in the same way as in Table 2. There were really 13 trait-pairs on the rating sheet, but we omit the last one, which dealt with strength of sex drive. The judges seldom used this category, and the results are not important for our problem.

For convenience we shall give each picture a number corresponding to the number of the set it is grouped with in the Szondi Test. On the test cards this number is actually a roman numeral. The judges looked at each picture by itself. We showed them *k1* first, then *p1*, *k2*, *p2*, *k3*, *p3*, and so on. They did not know what kind of picture they were rating.

SUBJECTS

Our subjects were either college students or college graduates. They are not a sample of people in general, but they are probably typical of intelligent adults. Most of them were a little older than the usual college group. Table 1 gives more information about the 40 judges and the 28 subjects who took the standard Szondi Test. The tested group included four Negroes; the five college graduates in this group were all honor graduates. The two groups are similar, but the Q1 (percentile 25) and Q3 (percentile 75) values for age show that the test subjects were slightly older and more heterogeneous in this respect than were the judges.

TABLE 1
DESCRIPTIVE DATA FOR THE GROUP OF JUDGES AND THE GROUP OF TEST SUBJECTS

	N	Per Cent Male	Per Cent Single	Per Cent Graduates	Age Q1	Age Q3
Judges	40	62	70	0	19.5	23.0
Subjects	28	61	68	18	22.3	29.5

RESULTS

Traits Attributed to Pictures. We tabulated the number of times each picture was judged to show each trait. Then we found what proportion of the total votes for each picture was given to the first member of the trait-pair. For example, 21 judges voted on picture *p2* when rating the trait-pair "ambitious-easygoing." Eighteen, or 85.6 per cent, of these judges marked "ambitious." If we assume that the two members of the trait-pair represent opposite ends of a single trait continuum, then this 85.6 per cent can be used as a score to show how clearly this picture represents "ambition." In the case of picture *p4*, 100 per cent of the 26 judges who rated it on this trait judged it to be "ambitious." Presumably *p4* more clearly produces the impression of an ambitious person than does picture *p2*.

After this kind of computation was carried out for all 12 traits, it was a simple matter to calculate the mean scores for the six *k* pictures and compare them to the corresponding means for the *p* group. The means for every trait are shown in Table 2. The significances of the differences were computed by applying the *t* test appropriate to small samples. Not a single difference yielded a *t* value large enough to be significant at the 10 per cent level. The largest difference is for the characteristic "wants tender care." The difference of 20.7 between *k* and *p* means yields a *t* of only 1.69. With 10 degrees of freedom, this value is well below that needed for the 10 per cent level of significance.

We are forced to conclude that *k* and *p* pictures do not form clearly differentiated groups so far as the kinds of traits attributed to them are concerned.

We can make another approach to comparing *k* and *p* pictures by listing the traits attributed to the pictures by a majority of all our judges. In Table 3 are shown the traits

TABLE 2
MEAN SCORES OF *k* AND *p* PICTURES FOR EACH TRAIT

Trait	*p* Mean	*k* Mean	Difference
Ambitious, hardworking Easygoing, satisfied	69.0	62.6	6.4
Likes to be with people Likes to be alone	39.1	44.1	5.0
Dominates, bosses Submits, gives in	56.2	49.4	6.8
Feels confident Feels inferior	48.4	39.1	9.3
Kind, helpful Cruel, malicious	48.9	49.2	0.3
Calm, relaxed Tense, easily upset	43.6	37.1	6.5
Jolly, fun-loving Sober, serious	15.7	33.7	18.0
Thinker, dreamer Practical, factual	43.4	53.9	10.5
Self-centered Interested in others	69.7	61.3	8.4
Optimistic, cheerful Pessimistic, cynical	24.9	41.3	16.4
Wants tender care Rejects tender care	41.3	62.0	20.7
Desires emotional ties with others Wants freedom, independence	42.1	60.3	18.2

TABLE 3
TRAITS ATTRIBUTED TO EACH PICTURE BY 50 PER CENT OR MORE OF FORTY JUDGES

Picture	Traits
p1	Feels inferior (65);* likes to be alone (50)
p2	Dominates (72); confident (57); tense (52); self-centered (52)
p3	Kind (60)
p4	Confident (75); dominates (70); ambitious (65); sober (57); self-centered (52)
p5	Thinker (60); likes to be alone (55)
p6	Feels inferior (52)
k1	Tense (70); feels inferior (62); likes to be alone (60); sober (52)
k2	Submits (60); wants tender care (57); feels inferior (50)
k3	None
k4	Confident (65); ambitious (52)
k5	Tense (57); feels inferior (52)
k6	Likes to be with people (72); jolly (65); easygoing (62); confident (50)

*Numbers in parentheses are percentages of judges (rounded to nearest per cent) marking the trait.

attributed to each picture by 20 or more of the 40 judges. Inspection of these data indicates that *k* pictures seem a little more likely to be judged as "tense" or as "feeling inferior." Two *p* pictures express dominance, but none of the *k* pictures appear to show this trait. These differences are not large, however, and provide little justification for believing that *k* pictures have common traits that distinguish them from *p* pictures.

Is it fair to expect differences between the two kinds of pictures using the traits on our list? We think it is. Deri [1, p. 187] suggests that the *k* pictures express certain features of catatonic schizophrenics such as a "minimum of fluid emotional contact between libidinal energy and the evironment," and "generally exaggerated tendency for seclusion and lack of contact with persons in the environment." The *p* factor, however, "refers to communication and contact with the outside" [1, p. 173]. At least three traits on our list should have characterized *k* pictures, according to this view: "likes to be alone," "self-centered," and "wants freedom, independence." But we have already shown that differences between *k* and *p* pictures are minor and not significant.

Grouping according to Facial Tension. After discovering that grouping according to diagnosis bore no relationship to the traits attributed to the pictures, we thought it would be rewarding to group the pictures according to certain facial cues. Thornton [4] and Samuels [3] have both shown the importance of various facial cues and of accessories such as glasses in affecting judgments. We chose to study the effect of muscular contraction in the face.

Ten graduate students, naive about the Szondi Test and the purpose of this study, were asked to classify the 12 ego-vector pictures into three groups: smiling, contracted, and relaxed. The contracted group was the most important. The students were given this definition of contraction: ". . . pictures which show *active muscular contraction* of some part of the face (except for smiling). Do not confuse wrinkles (which do not show present active contraction) with the furrows caused by contraction. Wrinkled faces may be contracted or relaxed." The relaxed group included the pictures where the face was in repose.

When we arrange the pictures according to the number of times each was judged to show muscular contraction, two clear sets of pictures can be formed. Pictures *k3*, *p2*, and *p6* all received ten placements in the contracted group. The other pictures frequently judged in this group were *k1* (nine placements) and *p5* (seven placements). All the other pictures received three or fewer placements in the contracted group. We thus have five pictures in the High Contraction group and seven in the Low Contraction group.

These groups can be analyzed for differences in the traits attributed to them in exactly the same way we used for the *k* and *p* groups. Results are shown in Table 4.

TABLE 4
MEAN SCORES OF HIGH CONTRACTION AND LOW CONTRACTION PICTURES FOR EACH TRAIT

Trait	High Mean	Low Mean	Difference	t*
Ambitious, hardworking	61.2	69.1	7.9	—
Likes to be with people	29.3	50.4	21.1	—
Dominates, bosses	59.2	48.2	11.0	—
Feels confident	36.6	48.8	12.2	—
Kind, helpful	26.0	65.6	39.6	3.50
Calm, relaxed	25.1	51.2	26.1	—
Jolly, fun-loving	13.6	32.6	19.0	—
Thinker, dreamer	50.0	47.7	2.3	—
Self-centered	87.3	49.9	37.4	3.38
Optimistic, cheerful	12.3	48.0	35.7	2.86
Wants tender care	37.5	61.8	24.8	—
Desires emotional ties with others	39.1	59.8	20.7	—

*Only significant values of t are listed. For 10 degrees of freedom, P = .05 when t = 2.23; and P = .01 when t = 3.17.

This table presents a very different picture from Table 2. In Table 4 differences of 20 points or more between the two means are common, and three of the differences are significant at better than the 2 per cent level. The pictures showing active muscular contraction are less often judged as kind and optimistic, and more often judged as self-centered, than are those without such contraction.

It is quite clear that groupings based on facial characteristics are more meaningful

than those based on diagnosis of the persons photographed. And it should be stressed that our division of the pictures on the basis of contraction is a very crude one. All the modifying effects of head posture, direction of eyes, presence of glasses, lighting, and position of arm and hand are neglected. Finer differentiations of facial cues could be made, and more intricate patterns of relationships could easily be discovered. But our point is sufficiently established by this one kind of analysis.

Ranking Pictures on Each Trait. Using the scores assigned to each picture, we ranked the 12 pictures from the highest to the lowest scores on each trait continuum. There were 12 rankings of the pictures—one for each trait. Thus *k6* earned rank 12 for "ambition," rank 1 for "likes people," rank 3 for "confidence," rank 10 for "dominates," and rank 1 for "jolly." The toothy smile of this picture probably has more to do with these impressions than the fact that the subject is catatonic.

Determining Preference Values. To measure the attractiveness of each picture in the test situation, we used the data from the 28 standard administrations of the Szondi Test. We tallied the number of times each picture was chosen as liked and the number of times it was chosen as disliked. Using the sum of these choices as a base, we computed the percentage of times it was liked. Thus one person chose *k1* as liked, and 20 people picked it as disliked. The preference value is 1/21 or 4.8 per cent. These values allowed us to rank the pictures from most liked to least liked as follows: *p4, p5, p3, k2, k6, p2, p1, k4, k5, p6, k1, k3.* Preference values range from 91.6 to zero. The *p* pictures are somewhat more popular than the *k* pictures.

M. R. Harrower [2] presents data on the choices of Szondi pictures by 100 normal males, 100 poorly adjusted males, and 115 poorly adjusted females. We have computed preference values for each ego-vector picture on the basis of her data. These values ranged from 93.7 for *p3* to 0.9 for *k1.* They yield the following rank order: *p3, p4, k6, p2, k2, p1, k4, p5, k5, p6, k3, k1.* When this order is correlated with that based on our own 28 subjects, *rho* equals .81. The most striking difference between the two orders is the greater popularity of *p5* (the thinker) among our subjects. The agreement between the two orders is remarkable, since the subjects tested in the two studies were from different populations. Another difference is that Harrower used a group-testing method with some of her subjects.

Correlations between Trait Rankings and Preference Rankings. With preference ranking as a criterion, we ran rank-order correlations between it and each of the 12 trait rankings. The *rho* values in Table 5 show how the various traits attributed to the pictures are associated with choice or rejection.

TABLE 5

RANK-ORDER CORRELATIONS BETWEEN RANKING OF PICTURES FOR TRAITS AND FOR PREFERENCE

Traits	Correlation with Preference	
	This Study	Harrower's Data
Calm vs. tense	.74	.70*
Kind vs. cruel	.65	.73
Confident vs. feels inferior	.50	.53
Likes to be with people vs. likes to be alone	.50	.64
Interested in others vs. self-centered	.38	.59
Jolly vs. sober	.38	.35
Optimistic vs. pessimistic	.28	.60
Wants tender care vs. rejects tender care	.22	.20
Desires emotional ties vs. wants freedom	.15	.22
Ambitious vs. easygoing	.05	.04
Thinker vs. practical	—.04	—.27
Dominant vs. submissive	—.17	—.11

* Values higher than .576 are significantly greater than zero at the 5 per cent level.

A fair summary of Table 5 is that subjects taking the Szondi Test definitely prefer pictures of people who appear kind, calm, and sociable to pictures of those who seem cruel, tense, and seclusive. Degree of ambition, thoughtfulness, dominance, or need for tender care have little relation to preference. These 12 pictures could be used to get at the degree to which subjects valued the appearance of kindly sociability in others, but the pictures would have to be regrouped without reference to diagnosis to accomplish this.

DISCUSSION

Present theories about the Szondi Test assume that the determinants of choice are re-

lated to the dynamics of the disorders of the photographed patients. We do not believe that this assumption squares with the facts. If you ask judges to tell what traits the pictures show, the paranoid and catatonic pictures are not clearly different; if the pictures are inspected for facial expressions and gestural cues, the *k* and *p* pictures are not clearly different. On the other hand, when pictures are grouped on the crude basis of degree of muscular contraction in the face or on the basis of the possession of certain socially desirable traits, differences between these groups can easily be found.

Our findings require another theoretical approach to the test. Here is one possible and plausible formulation: Through long acquaintance with other human beings, a person has learned to react rapidly and fairly accurately to the faces around him. Frowns, smiles, and fleeting expressions of annoyance have come to signal his own acceptance or rejection by others. Tense people have proved to be unpleasant companions, while cruel words have rarely come from relaxed or smiling faces. Thus, people acquire tendencies to approach or avoid certain facial expressions. Faced with a series of strange-looking pictures, these approach or avoidance tendencies operate to produce choice or rejection.

We think that the Szondi principle of using choice of pictures to reflect personality trends is a useful one when viewed in this light. People who have grown up in meagre or distorting social atmospheres probably have acquired peculiar sets of approach-avoidance patterns to human faces. Their choices of pictures should reflect these peculiarities. With an adequate understanding of the stimulus value of the pictures and a grouping based on something more meaningful than diagnostic categories, we could very likely make some correct inferences about their personalities.

SUMMARY

Forty judges rated the 12 pictures from the ego-vector of the Szondi Test on 12 traits. These ratings allowed each picture to be assigned a score on each trait. When the catatonic and paranoid pictures were compared, no important differences in the traits assigned to each diagnostic type could be found. When the pictures were grouped on the basis of muscular contraction in the face, several significant differences appeared in the traits assigned to the groups with differing degrees of contraction.

Data from the standard administration of the Szondi Test permitted ranking of the pictures for degree of preference. Preferred pictures tend to be those judged kind, calm, and sociable. Ambition, dominance, thoughtfulness, and need for tender care are not related to preference.

These findings suggest that test subjects choose pictures which appear accepting and friendly. They reject pictures that appear cruel and asocial. Probably approach and avoidance tendencies acquired in actual social interaction determine choice in the test situation.

Our results do not support the view that the dynamics of choice are related to the disorders of the photographed patients. They are consistent with view that choices depend upon learned reactions to facial expressions.

Received September 12, 1950.

REFERENCES

1. DERI, SUSAN. *Introduction to the Szondi Test.* New York: Grune & Stratton, 1949.
2. HARROWER, MOLLY R. Experimental studies with the Szondi Test. Psychological Service Center, Univ. Okla., 1949. (Mimeo.)
3. SAMUELS, M. R. Judgments of faces. *Character & Pers.*, 1939, 8, 18-27.
4. THORNTON, G. R. The effect of wearing glasses upon judgments of personality traits of persons seen briefly. *J. appl. Psychol.*, 1944, 28, 203-207.

COUNSELING METHOD AND CLIENT RESPONSES[1]

DANIEL V. BERGMAN
UNIVERSITY OF CHICAGO

THE relationship between the ways in which the counselor responds to the client during nondirective interviews and a predictable process of therapy has been a source of thoretical discussion [8] and of experimental study [1, 3, 7, 9, 10]. Certain studies have explored the total process of nondirective therapy in terms of the types of responses made by the client and counselor [9, 10]. A study by Porter [5, 6] provided a reliable basis for classification of the types of counselor responses. Hogan [2] furthered the work of arriving at objective measures of client attitudes.

The present study examines a portion of the process of nondirective therapy in terms of three interrelated categories based on the recorded verbal behavior of client and counselor. These categories are designed to describe the types of client and counselor responses which occur immediately after the client has requested an evaluation from the counselor. The writer selected this phase of the counseling process in the belief that if a systematic relationship exists between counseling method and client responses, its nature, maximum variation, and relevance to the validity of the nondirective hypothesis would be most explicitly revealed by the interaction of client and counselor during such portions of the interviews. No attempt is made to explore other phases of the process or the total verbal behavior of client and counselor.

PROCEDURE

The stages in the present study were: (1) procurement of the research case material, (2) development of categories of analysis, (3) application of these categories to the data, (4) testing the reliability of this classification, and (5) analysis and interpretation of the findings. Each stage will be described below.

The Data. The data consist of the frequencies of occurrence of response units described below. These units have been assigned equal statistical weight irrespective of the cases from which they are drawn. All data were taken from electrically recorded interviews. The interviews used were from a block of ten counseling cases made available for research by the Counseling Center of the University of Chicago. These cases represent the work of several counselors trained in the nondirective method.[2]

The Response Unit. A response unit consists of (1) the client's total verbal expression in which a request is made for a counselor evaluation, (2) the total verbal expression of the counselor immediately following presentation of this request, and (3) the total verbal expression of the client immediately following this counselor response. The passages below illustrate a response unit:

> Client: That's right. Uh huh. That's very true. I suppose it all fits in with a lack of self-confidence and, therefore, I mean, a lack of confidence in what I am and, therefore, a need to hide what I really am and to adapt in terms of what I'm supposed to be. I still don't see any way of changing. What does one do about it?
>
> Counselor: You are saying there that, uh, even though one can realize some of those things, still it's a little bit hard to know how to change.
>
> Client: Uh huh. I suppose it's essentially a matter of building up confidence in myself so that I wouldn't have these feelings of inadequacy and, therefore, the need to escape from them either by running away into a field where I don't have them as I'm doing now or else to retaliate on the person who arouses

[1]This report is based upon research submitted by the writer in a Master's paper to the Committee on Human Development at the University of Chicago.

[2]For a complete description of the characteristics of the cases which make up this "research block," see Nathaniel J. Raskin, "The Development of the Parallel Studies Project," *J. consult. Psychol.*, 1949, 13, 154-156.

them more strongly. I guess that really is it. When I have been thinking about this whole marriage situation, I've always thought of it in terms of, oh, for about a year anyhow, I think I realized that I wasn't getting full satisfaction out of it and I was, when I first started thinking I thought well I'll just have to accept the fact that I can't. I'll have to find some outside way of satisfying my needs or interests or whatever they are. And I suppose essentially the problem is of getting satisfaction within myself so I won't need this external thing.

For convenience, the three elements of the response unit are coded and abbreviated below. Each of the three elements was classified in terms of several categories. These categories, each illustrated with data used in the present study, are presented below:

A. *Client request for evaluation.*

1. *Client requests solution of his adjustment problems.*

Uh huh. I hate to do it, but I do it anyway. [Pause] And so I just, I don't know where to start in trying to look for something. I think what I'd really like would be for somebody to just sit me down and give me a series of objective tests or something and say: "This is the kind of person you are and this is what you should do and here is the formula for happiness." Now [laugh] and I'd just do it. [Pause]

2. *Client requests evaluation of adjustment or of progress toward adjustment.*

Yeah and that is one question I was going to ask you. You've had experience along this type in here or else you wouldn't be here. It didn't, my talk and all didn't have a basic neurotic background to it, to you, did they?

Do you think this has helped me today? Is this more what I — you — oh — expect as the outcome, or what?

3. *Client requests confirmation of a viewpoint, projected choice, or projected action.*

Yeah, uh huh, I mean, can you see the difference from this type and this?

Tell me if it's just hopeless and I am just at an age that I do have to sit and wait while life passes me by — maybe.

So shall I burn up the notes and not take the exams? Jump out the window or — ? [Laughs] Or should I stay up all night with black coffee and wet towels and cram it in?

4. *Client requests selection of fruitful types of activity.*

Well, I have my afternoons free, so that, so just at any time. If you could just give me some hint of what the general procedure is, I probably could carry it out, but I, uh, I don't quite know what to do, what to start.

B. *Counselor response to the request.*

1. *Counselor interposes his own viewpoint by agreement or disagreement, suggesting a possibility for further exploration, supplying information, or presenting conclusions about the material presented by the client.* (This category includes those responses frequently classified as "interpretation.")

Well, in other words, the relationship with this, ah—This doesn't seem to be genuinely human. Doesn't seem to be exactly real. And I don't think you feel that it's really helpful either.

As I — as far as I know, I expect that it means the—well, just as we speak of a — an engine might run, but one wants it to run efficiently — does it do its job? Does it function up to its full horsepower? I guess that's the term.

— or a third alternative would be to keep both of them on, and make some kind of adjustment, uh, on each so they work in harmony.

2. *Counselor states a theory of therapy, his view of his own role or of the client's role, but refrains from any suggestion to the client beyond the possibility of success from continued exploration.* (This category includes those responses generally referred to as "structuring.")

Well, I think that the thing we've found is that a person really gets the farthest if they explore their own situations and I don't know, uh, I don't know if you wish to do that but if you do, why you could talk about any angle of it that — that you'd been thinking about or —

3. *Counselor indicates that the meaning of the client is not clear to him by asking for clarification or by presenting a reflection in tentative terms.* (This category parallels the familiar "request for clarification" classification.)

You feel, really, that you'd like me to try to say something. And I suppose that the one thing that I could say, I don't even know that you'll believe. That is, that I'm trying to understand how this whole situation feels to you because I think as you explore your way through it, that as I understand what it is you're exploring through, then we could arrive. But when you ask, "What should I do," I suppose you're asking about exams, primarily, or at least partially. Is that part of it?

4. *Counselor reflects one or more feelings inherent in the response encompassing the request for evaluation but does not recognize verbally the request element itself.*

Client: But now how can I feel adequate to make a decision? I mean, how can I tell which is best, which I would be happier doing and which would be best for us? I mean I sort of have a tendency to feel all alone — that it would be better for me, but that really isn't the kind of life that I want — marriage and all of the responsibilities that go with it.

Counselor: You're saying there that you lean a little bit toward a kind of life that would lead away from marriage —

5. *Counselor presents his understanding of the feeling represented by the request for evaluation with or without reflection of feelings embodied in other elements of the encompassing response.* (Both 4 and 5 include responses usually classified as "reflection of feelings.")

Client: If I start to work on a problem, and if I definitely decide to work on a course of action, then the action, more or less, will be completed. [Pause] And that seems to be just about it. And so I was wondering whether you had any specific advice or if there was anything that you'd like to advise me to do on this specific problem.

Counselor: You just wonder if somebody else can look at this thing a little more from a distance and see some kind of an answer there.

C. *Subsequent client expression.*

1. *Client again presents a request for evaluation which may be the same as the original or may be modified by lessening or increasing the scope of the attitude or problem presented for evaluation.*

Client: I mean, doesn't it seem that way to you? I wish you'd tell me really, even though — and something else I thought of. I think that I really don't, that I don't really believe for myself in nondirective counseling. I mean I must not. I must have these other ideas because I keep thinking, well, some things must be so much clearer to you and if only you'd tell me. I think I'm in a receptive mood and I think I'll accept things.

Counselor: You come to the feeling occasionally that after my hearing what you say in this situation that I must know more than you, is that it?

Client: Sure. It must be more apparent to somebody else. And I think I want to change things enough so that — I mean, that I'll accept an explanation from somebody else. Or at least think about it.

2. *Client accepts or rejects counselor response and abandons the attempt to explore his attitudes and problems without again presenting a request for evaluation.*

Or else a life-saver so I don't drown, instead of an anchor. [Weeping; long pause] Guess there's nothing more I can say. Just go home and take a sleeping pill — sleep it off.

3. *Client continues to explore his attitudes and problems.*

Counselor: That kind of behavior seems incompatible with being a social worker, to you.

Client: Yeah. I mean, of course that doesn't presuppose that social workers are, uh, witless, and uh, oh, lack any form of humorous expression, and all. But I mean, basically, they should be, oh, mature looking and professional and businesslike, and, oh, you can laugh, but you should have a sense of humor and all that — I mean my behavior exceeds the amount of humor that would be accepted.

4. *Client verbalizes an understanding of relationships between feelings presented or between his attitudes and his behavior in counseling or in the life-situation.*

Counselor: You feel that it's hard for you to be honest with yourself and that it's sometimes easy for you to entertain the thought of letting some outside person do it because he can be more objective.

Client: M-hm. Mhm. Yes. It also would be much simpler than — I mean it would mean less responsibility.

In the analysis of the data which follows, the categories elaborated above will be referred to more frequently by the following abbreviations:

A. Client request for evaluation.
  A1. Solution of problems.
  A2. Adjustment or progress.
  A3. Confirmation of viewpoint, choice, or action.
  A4. Leads for exploration of problems.
B. Counselor response to the request.
  B1. "Interpretation."
  B2. Structuring.
  B3. Request for clarification.
  B4. Reflection of feeling: prior-post.
  B5. Reflection of feeling.
C. Subsequent client expression.
  C1. Repetition of request.
  C2. Abandonment of self-exploration.
  C3. Continued self-exploration.
  C4. Insight.

Application of the Categories. The categories of the scale presented above were applied to 246 response units selected from 60 interviews which constituted 10 counseling cases. The number of interviews per case ranged from three to nine. (Follow-up interviews were not included in the study.) The writer's selection of response units per case ranged from zero to 65.

The scale was designed to yield a numerical score for each category and for each relationship which could exist between categories. Thus, the writer judged each response unit in regard to (A) the type of request for evaluation, (B) the type of counselor response to the request, and (C) the type of the client's subsequent expression. Each response unit was then assigned a frequency of one which expressed the particular combination of the categories (A), (B) and (C) into which its three elements were classified.

Reliability of the Classification. Porter [5, 6] and Snyder [10] tested the reliability of judgments in regard to data selected from

nondirective interviews by having the same judges (themselves) rescore the same material after a period of time had elapsed. Porter also trained other judges and compared his own judgments with those which they achieved. Curran [1] employed both comparison of agreement between judgments made by the same judge and agreement between those made by separate judges. The findings of these studies indicate that when judges are trained in the categorization of material from nondirective interviews, their ability to repeat the same distinctions does not substantially differ from the ability of different judges to make the same distinctions once.

The writer, therefore, assumed that if he had trained other judges than himself they would have responded to the data of the present study in the same way that he responded. The writer accordingly scored all response units in the manner already described. After a period of three months had elapsed, each response unit was rescored in the same manner. Where both scores were the same, the frequency was listed under *agreement*. Where the scores failed to coincide, a tally was placed under the category of the original score together with a notation of how it was rescored. (Thus, "A1 rescored as A3," etc.) The frequencies of the noncoincident scores were combined for each category under *disagreement*.

The results of these applications of the scale categories to the response units have been summarized in Table 1. In this table, *degree of agreement* is the product of the frequency of agreement over the total frequency for each category. The meaning of this measure of reliability may be obtained by comparison with the final column. This column lists the degree of agreement required when P is equal to .00135. This figure was obtained by a restatement of the Test Ratio suggested by Peatman [4, p. 375].

Inspection of Table 1 reveals that the degree of agreement obtained for ten of the 13 categories is significant at the 0.1 per cent level of confidence. The degree of agreement for all categories grouped under each element of the response unit is: client request, 85.8; counselor response, 82.5; and subsequent client expression, 79.3. A degree of agreement of 59.5 in each case is significant at the 0.1 per cent level of confidence. This indicates that the writer was able to assign the same meaning to the verbal expressions of client and counselor with a high degree of reliability after a time lapse of three months.

TABLE 1
FREQUENCY OF USE OF EACH CATEGORY AND THE RELIABILITY OF THE CLASSIFICATION

Category	Frequency	Degree of Agreement	Degree of Agreement required if P = .00135
A1. Solution of problems	106	80.2	64.6
A2. Adjustment or progress	88	90.9	66.0
A3. Confirmation of viewpoint, choice, or action	29	79.7	77.9
A4. Leads for exploration of problems	23	91.3	81.3
B1. Interpretation	39	94.9	75.8
B2. Structuring	27	66.7*	78.5
B3. Request for clarification	35	62.1*	75.4
B4. Reflection of feeling: prior-post	22	59.1*	80.5
B5. Reflection of feeling	123	91.8	63.5
C1. Repetition of request	78	87.2	67.0
C2. Abandonment of self-exploration	61	75.4	69.2
C3. Continued self-exploration	78	68.2	67.0
C4. Insight	29	96.6	77.8

*Not significant at the .001 level of confidence.

ANALYSIS OF THE DATA

The purpose of the classification of the data reported above was to reveal the influence of the therapeutic relationship upon client and counselor activity. The analysis of the categorized data was designed to answer two central and six subsidiary questions. These questions were:

1. Do systematic relationships exist between requests for evaluation, counselor responses, and subsequent client expressions?

a) What are the relationships between counselor responses and subsequent client expressions?

b) What are the relationships between requests for evaluation and subsequent client expressions?

c) What are the relationships between requests for evaluation and counselor responses?

2. What is the nature of the relationship between counseling method and client self-exploration and insight?

a) What types of counselor response are followed more often by self-exploration and insight?

b) What types of counselor response are fol-

lowed more often by repetition of the request?

c) What types of counselor response are followed more often by abandonment of self-exploration?

The Systematic Relationships. The scale employed in the study provides for the possibility of 80 types of relationships between the three elements of the response unit. Actually, 49 different combinations occurred with frequencies ranging from 1 to 28. In order to determine the statistical significance of these observed frequencies, the probability of occurrence of any relationship (.0125) was placed in the restatement of the Test Ratio employed earlier. It was determined that any such combination occurring .0338 times was equal to a probability of .00135. Nine of the relationships fulfilled this criterion, ranging in frequency of occurrence from .0376 to .1178. The significant relationships are presented in Table 2.

TABLE 2

FREQUENCY OF EACH RELATIONSHIP SIGNIFICANT AT THE .001 LEVEL OF CONFIDENCE*

Categories Related	Frequency of Occurrence	Percentage of Occurrence
A1 : B1 : C2	9	3.76
A1 : B3 : C1	11	4.57
A1 : B5 : C1	12	5.38
A1 : B5 : C3	21	8.84
A1 : B5 : C4	16	6.70
A2 : B3 : C1	12	5.38
A2 : B5 : C3	28	11.78
A3 : B1 : C2	11	4.57
A3 : B5 : C1	12	5.38

*For description of categories see Table 1.

The Relationships between Counselor Responses and Subsequent Client Expressions. The findings presented in Table 2 reveal systematic patterns of significance. In order to explore the uniformity of such patterns in terms of the total data, a chi-square analysis

TABLE 3

FREQUENCIES OF VARIOUS CLIENT EXPRESSIONS WHICH FOLLOW WHEN COUNSELOR RESPONDS IN DIFFERENT WAYS TO REQUESTS FOR EVALUATION

Counselor Responds By:		Client Subsequently: C1 Repeats Request	C2 Abandons Self-exploration	C3 & C4 Continues Self-exploration or Achieves Insight	Total
B1	*o*	5	27	7	39
	e	12	10	17	
Interpreting	χ^2	−4.08	+28.90	−5.88	38.86
	P	.05	.001	.02	.001
					.71
B2	*o*	9	14	4	27
	e	9	6	12	
Structuring	χ^2	–	+10.67	−5.33	16.00
	P	–	.01	.05	.001
					.61
B3	*o*	24	5	6	35
	e	11	9	15	
Requesting	χ^2	+15.36	−1.78	−5.40	22.54
clarification	*P*	.001	–	.05	.001
					.63
B4 & B5	*o*	40	15	90	145
	e	46	35	64	
Reflecting	χ^2	−0.78	−11.43	+10.56	22.77
feeling	*P*	–	.001	.01	.001
					.37
Total		78	61	107	246
χ^2		20.22	52.78	27.17	100.17
$P(\chi^2)$		.001	.001	.001	.001
C		.45	.68	.45	.54

TABLE 4
FREQUENCIES OF VARIOUS CLIENT EXPRESSIONS WHICH FOLLOW THE ORIGINAL REQUESTS FOR EVALUATION

Client Requests:		Client Subsequently: C1 Repeats Request	C2 Abandons Self-exploration	C3 & C4 Continues Self-exploration or Achieves Insight	Total χ^2 $P(\chi^2)$
A1	*o*	35	20	51	106
	e	34	26	46	
Solution of problems	χ^2	+0.03	−1.38	+0.54	1.95
	P	–	–	–	–
A2	*o*	32	21	35	88
	e	28	22	38	
Evaluation of adjustment or progress	χ^2	+0.57	−0.05	−0.24	0.86
	P	–	–	–	–
A3	*o*	5	13	11	29
	e	9	7	13	
Confirmation of viewpoint, choice, or action	χ^2	−1.78	+5.14	−0.31	7.23
	P	–	.05	–	.05
A4	*o*	6	7	10	23
	e	7	6	10	
Leads for exploration of problems	χ^2	−0.14	+0.17	–	0.31
	P	–	–	–	–
Total		78	61	107	246
χ^2		2.52	6.74	1.09	10.35
$P(\chi^2)$		–	–	–	–

was made of all combinations of counselor responses and subsequent client expressions which occurred.[3] The results are presented in Table 3. The plus and minus signs indicate the direction of the deviations of observed frequencies from chance expectancy. Contingency coefficients were derived from each chi-square value which proved to be statistically significant. The chi-square value for this analysis (100.17) is significant at the 0.1 per cent level of probability. The contingency coefficient (.54) indicates a moderate correlation and substantial relationship between counselor responses and subsequent client expressions.

The Relationships between Requests for Evaluation and Subsequent Client Expressions. The findings embodied in Table 2 failed to suggest the existence of systematic relationships between requests for evaluation and subsequent client expressions. To discover whether such relationships existed elsewhere in the data, a chi-square analysis was made of all combinations of these two elements. The results are presented in Table 4.

The distribution of client expressions which follows the requests for evaluation presents no evidence of a relationship between the type of request and the type of subsequent expression. The chi-square value for this analysis (10.35) is without statistical significance. However, one cell-value expresses a deviation significant at the 5 per cent level of confidence. At this level, a relationship is indicated between re-

[3]It will be noted that two categories each of counselor response and subsequent client expression have been grouped together in this analysis. This was done to eliminate cells in which expected frequencies would be less than five. Since both counselor categories are reflection of feeling and both client categories represent behaviors invariably considered to be positive in therapeutic terms, the writer does not consider that the meaning of the results is altered by such grouping.

quest for evaluation of a viewpoint, choice, or action, and abandonment of self-exploration. The over-all value of the table deviation and the conformity of 11 of the 12 cells to this value seem to indicate that the relationships which have been revealed are a product of client and counselor interaction.

The Relationships between Requests for Evaluation and Counselor Responses. Relationships between the type of request and the type of counselor response are not indicated by the findings presented in Table 2. The exploration summarized in Table 3 revealed highly significant relationships between the type of counselor response and the type of subsequent client expression. That these relationships were a product of client and counselor interaction rather than of the client's independent "set" was substantiated by the findings given in Table 4. Whether the counselor responses were independent of the type of request remained to be determined by a chi-square analysis of all combinations of requests and counselor responses. The results are presented in Table 5.

The writer's conclusion that the relationships discovered are a product of client and counselor interaction is strongly re-enforced by the findings embodied in Table 5. The chi-square value for all relationships (40.32) is significant at the 0.1 per cent level but correlation is low (.38). However, the correlation between request for confirmation of viewpoint, choice, or action and leads for exploration of problem with all counselor responses, and between interpretation and all client requests is high. In both cases the major contribution to variance is made by the relationship between the client and counselor categories named. Since both of these categories have previously been found to be significantly related to abandonment of self-exploration,

TABLE 5

FREQUENCIES OF DIFFERENT WAYS IN WHICH THE COUNSELOR RESPONDS TO VARIOUS REQUESTS FOR EVALUATION

Client Requests:		Counselor Responds By: B1 Interpreting	B2 Structuring	B3 Requesting Clarification	B4 & B5 Reflecting Feeling	Total
A1	*o*	12	8	14	72	106
	e	17	12	15	62	
Solution of problems	χ^2	−1.47	−1.33	−0.07	+1.61	4.48
	P	–	–	–	–	–
						–
A2	*o*	7	9	19	53	88
	e	14	10	12	52	
Evaluation of adjustment or progress	χ^2	−3.50	−0.10	+4.08	+0.02	7.70
	P	–	–	.05	–	–
						–
A3 & A4	*o*	20	10	2	20	52
	e	8	6	7	31	
Confirmation of viewpoint, choice or action &	χ^2	+18.00	+2.67	−3.57	−3.90	28.14
	P	.001	–	–	.05	.001
						.59
Leads for exploration of problems						
Total		39	27	35	145	246
χ^2		22.97	4.10	7.72	5.53	40.32
$P(\chi^2)$		.001	–	.05	–	.001
C		.61	–	–	–	.38

the seeming nonconformance of one cell of Table 4 to the general finding seems adequately explained.

The Nature of the Relationship between Counseling Method and Client Self-exploration and Insight. The chi-square analyses presented established the existence of systematic relationships between counseling method and subsequent client expressions. These relationships were revealed as concomitants of client and counselor interaction. The specific nature of the relationships which exist between the different types of counselor responses and the process of self-exploration and insight may be stated from the findings presented in Table 3.

Types of Counselor Response Followed More Often by Self-exploration and Insight. Reflection of feeling is the only counselor response which is followed more often by continued self-exploration or insight than would be expected by chance alone. The deviation is significant at the 1 per cent level. This counselor response is the only one which is followed less often by abandonment of self-exploration than would be expected by chance. The deviation is significant at the 0.1 per cent level.

Types of Counselor Response Followed More Often by Repetition of the Request. Request for clarification is the only counselor response followed more often by repetition of the request than would be expected by chance alone. The deviation is significant at the 0.1 per cent level. At the 5 per cent level, self-exploration and insight follow this response less often than would be expected by chance.

Types of Counselor Response Followed More Often by Abandonment of Self-exploration. Both "interpretation"[4] and structuring are followed more often by abandonment of self-exploration than would be expected by chance alone. The deviations are significant, respectively, at the 0.1 per cent and 1 per cent levels of confidence. At the 2 per cent level, "interpretation" is followed less often by self-exploration and insight than would be expected by chance. At the 5 per cent level, structuring is followed less often by self-exploration and insight, and "interpretation" less often by repetition of the request than would be expected by chance.

[4]The writer wishes to call attention at this point to the operational definitions of this and other categories supplied earlier in this report. It should be clear, for example, that "interpretation" as used in the present study is synonymous with any counselor verbalization which conveys meaning not present in the client's communication, such as information or an opinion. This definition does not of necessity conform to the usage of other psychological writers.

SUMMARY AND CONCLUSIONS

Hypothesis. The hypothesis to which the findings of the present study are relevant is that the predictable process of client-centered therapy follows when the therapist retains within his conduct of relationship only those procedures and techniques which convey his deep understanding of the emotionalized attitudes expressed and his acceptance of them and refrains from questioning, probing, blame, interpretation, advice, suggestion, persuasion, reassurance [8]. For the purposes of this study, the writer restated the hypothesis as a question: What is the nature of the relationship between counseling method and client self-exploration and insight?

Methodology. The hypothesis was rendered amenable to exploration in quantitative terms by the assumptions:

1. That this hypothesis can be tested by analysis of the verbal behavior of client and counselor during the course of nondirective, or client-centered, therapy.
2. That the relationship would be most clearly focussed at those phases of therapy in which the client had presented a request for evaluation, but that these phases do not differ in kind from other phases of therapy.
3. That a defined unit of response within a counseling interview can be treated as a statistical unit rather than as a context-determined particular of a unique whole.
4. That such response units can be assigned an equal statistical weight as representing the same qualitative factor as to meaning and relationship to process when derived from several cases which differ in terms of individuals involved, length, and "success."

These four assumptions have a common core in terms of previous research on therapy. It has become usual procedure in such studies to categorize and tabulate frequencies from the recorded verbal behavior of client and counselor. The use of this procedure in the pre-

sent study had certain aspects not present in such studies. The attempt was to reveal meaningful relationships residing in the verbal interaction of client and counselor. The response units were treated as interconnected and codetermining. The focus was interrelationship rather than discrete occurrence.

The writer believes that the findings presented demonstrate the truth of these assumptions. Whether the reader ascribes to the relationships discovered the meaning assigned to them by the writer, or not, their existence in numbers which transcend any doubt as to their statistical significance has been established. In terms of much research that is still needed, the establishment of these assumptions seems of importance. Although the purpose of this study is to present evidence in regard to an hypothesis as to the nature of the process of therapy, ultimately the importance of the conclusions in this regard may be less than that which resides in the discovery of a feasible way in which relationship and process may be studied by using concrete verbal behavior units.

Findings. Insofar as the sample obtained by the writer is both adequate in numbers and sufficiently random, insight into the relationship between counseling method and client self-exploration and insight has been gained from the analysis of the response units. If it is considered that the positive aspects of process derive from self-exploration and insight, that repetition of the request represents neither a positive nor a negative aspect of process, and that abandonment of self-exploration is negative to the achievement of process, the findings of the study may be summarized as follows:

1. Reflection of feeling is the only type of counselor response which is related in a significant manner to an increase in the frequency of positive aspects of the process and to a decrease in negative aspects of the process.
2. All other counselor responses studied are related, at various levels of significance, to a decrease in the frequency of positive aspects of the process.
3. "Interpretation" and structuring, as defined for this study, are related significantly to an increase in the frequency of negative aspects of the process.
4. Request for clarification tends largely to maintain the status quo in regard to the process although related to a decrease in frequency of positive aspects of the process.

Conclusions. Within the aspect of therapy explored by the present study and at the levels of confidence cited, self-exploration and insight are concomitants only of reflection of feeling, while abandonment of self-exploration is significantly related to "interpretation" and structuring. These findings are entirely consistent with the hypothesis which has been stated.

The writer does not consider that the findings of this study are conclusive proof of the validity of the hypothesis. It is believed, however, that insofar as the data employed constitute an adequate sample of the usual conduct of nondirective therapy, the evidence cited re-enforces the hypothesis. More important, the evidence indicates both the need and the potential value of research upon a larger scale and of a more encompassing nature upon the function of relationship in client-centered therapy as it bears upon process.

Received September 19, 1950.

REFERENCES

1. CURRAN, C. A. *Personality factors in counseling.* New York: Grune & Stratton, 1945.
2. HOGAN, R. A. The development of a measure of client defensiveness in the counseling relationship. Unpublished Doctor's dissertation, Univ. Chicago, 1948.
3. LEWIS, VIRGINIA. Changing the behavior of adolescent girls. *Arch. Psychol., N. Y.,* 1943, No. 279.
4. PEATMAN, J. G. *Descriptive and sampling statistics.* New York: Harper, 1947.
5. PORTER, E. H., JR. The development and evaluation of a measure of counseling interview procedures. I. The development. *Educ. psychol. Measmt,* 1943, 3, 105-126.
6. PORTER, E. H., JR. The development and evaluation of a measure of counseling interview procedures. II. The evaluation. *Educ. psychol. Measmt,* 1943, 3, 215-238.
7. RAIMY, V. C. The self-concept as a factor in counseling and personality organization. Unpublished Doctor's dissertation, Ohio State Univ., 1943.
8. ROGERS, C. R. Significant aspects of client-centered therapy. *Amer. Psychologist,* 1946, 1, 415-422.
9. SEEMAN, J. A study of the process of nondirective therapy. *J. consult. Psychol.,* 1949, 13, 157-168.
10. SNYDER, W. U. An investigation of the nature of nondirective psychotherapy. *J. gen. Psychol.,* 1945, 33, 193-223.

SITUATIONAL PLAY THERAPY WITH NORMAL CHILDREN

CLARK E. MOUSTAKAS
THE MERRILL-PALMER SCHOOL, DETROIT

IN the past two decades much clinical insight has been derived from studying the inner world of the disturbed child. Various theories and techniques of play therapy have been presented and evaluated by psychotherapists, and analytical and neoanalytical case studies, with complete interpretations, have appeared regularly and increasingly.

Client-centered recordings of nondirective play therapy sessions have dramatically portrayed the process of attitudinal reorganization in young children. Almost all the children described in these reports have had some degree of personal and social maladjustment, fixated psychosexual paths, or inadequate self-concepts.

The present study was concerned with normal children who were faced with a disturbing new family experience which they perceived as threatening to themselves, and who were given an opportunity to resolve their feelings in situational play therapy sessions. In the two cases here described in some detail both children were faced with one of the commonest crises of childhood, the arrival of a new baby in the family. Of a number of instances in which play therapy sessions were equally effective, these two were selected as providing interesting and clear-cut illustrations of the method.

THE NEW BABY CRISIS

Normal children who experience such catastrophes as fires and floods, or who have accidents or illnesses, or are subjected to such family crises as divorce and death, often show confusion, hostility, uncontrollable aggression, hate, and anxiety. The arrival of a new baby in the family is one of the commonest sources of such a disturbance in the child's behavior. To all children such an event brings a period of stress, for however stable, well organized, and rooted in positive emotions the family relationships may be, the arrival of a new family member requires some modification of role in each person in the family. Some family disorganization may result, at least temporarily, and the older child or children are often faced with a difficult adjustment to the new situation.

For neither of the two children, Tommy and Susan, whose play therapy sessions are reported was the arrival of a new baby a surprise. Both had been informed of the coming event two or three months in advance, and both had expressed pleasure in the prospect.

THE CASE OF TOMMY

Tommy, four years old, had been rated as fairly well adjusted both personally and socially by his nursery school teacher, the nursery school director, and the psychologist. His relations with other nursery school children were satisfactory, he came to school happily, and he talked with pride of his home and parents. His parents, in turn, talked of him with pleasure and pride and regarded him as a happy, secure, confident child who easily accepted limits and responsibilities.

When Tommy was four and one-half, an adopted girl of thirteen was suddenly brought into the home, and three months later his mother gave birth to a daughter. During this period Tommy's behavior showed a radical change, both at school and at home. At school he became sulky, refused to accept even simple, clear, and reasonable limits, showed a tendency to retreat from child groups whenever things did not go his way, and often withdrew into long sessions of solitary play.

At home he became fidgety at mealtimes, refused to eat foods that he had accepted before, cried, attempted to destroy the family record player, and often appeared ill-tempered and irritable. His mother, after attempting to deal with the situation by explanations and supports of many kinds, requested that he be given play therapy.

Three play sessions were conducted with Tommy. During the first he played with airplanes and trucks the entire time and was relatively quiet. In the second session Tommy appeared to focus on his attitudes about himself and what the two new members of the family would mean in terms of his role. He perceived them as potential threats, but once he had recognized these feelings and they had been accepted and clarified, he could accept his siblings and share with them his emotional and material possessions, and see his altered role as one that did not threaten his real self. A transcript of this session from the tape recording follows.

Second Play Session with Tommy

*T. You can use these things in any way that you want, Tommy.

†C. You know what? I could make a little castle out of that. [Indicates sand in sandbox.]

T. You could make a castle.

C. These are two boats. Look.

T. Mm-hm.

C. You know what kind? This one is a ship and this one is a ferry.

T. One's a ship and one's a ferry.

C. This is the ocean. [Points to sand.] This is the way that they use them in New Mexico.

T. In New Mexico, they use them like that.

C. Now do you know what we have to do? We have to get some water and smooth it. [Refers to sand.] You know what I can do? I can make an ocean liner and put it in the sand.

T. You can do that.

C. Then this can be the dock. [Points to hill of sand.] Then the ocean liner can go on it. It can go right on the hill.

T. Mm-hm.

C. There is just room for two boats to be on it. There. Now I'll make another boat. This can be the parking space. [Points to spot in the sand.] See? This is a great big parking lot for it.

T. A great big one.

C. See? This is where the little boat goes. He goes way up there. There's a parking lot for the big boat and one for the little boat. We have to do this over at the dock. Toot, toot, toot go the boats. Look where this boat has to go. He goes to get the sand. I'm putting it right in.

T. You're putting it right inside.

C. I'm pretending this is a ship. This is where they really go. Right over here. [Points to spot in sand.]

T. That's really the place for them to go.

C. This is the stuff that they carry into the dock. [Sand.] Look what he has to do. He's going to bury this whole big boat.

T. He'll bury the whole big boat.

C. See? I can bury him.

T. You're really burying him.

C. So no one will ever find him again.

T. He will be lost for good.

C. He'll be lost for good. It can't get out now. You see, this is the little boat's dock and no one ever goes in this dock. Because that's his dock.

T. Just his.

C. You know what he's going to do? He's going to put sand and water on this boat. [Points to big boat which he has taken out of sand.] Then he'll clean it up.

T. He'll make it clean.

C. Hi, Joe. Howdy. You see, this big one that comes along is his brother. You see, these two boats are brother boats.

T. One is the brother of the other one.

C. Yeah. One is the brother of the other one. Hey! Who messed my dock up? "Well, I did," says the big boat. See? He has some sand in him. He carries people in his boat, and this one has sand in his, too.

T. They both have sand.

C. You know what? They dump out the sand there. He [little boat] scrubs his boat off. Then he [big boat] scrubs his boat off. Both of them go. You know where they're going now? In the . . . in the brink the ship goes first [big boat]. Say, what do you know, Joe? I have to make another dock for this boat [a tiny boat].

T. Another dock for another boat.

C. Yeah. Oh, I know a nice dock. What do you know, Joe? He'll be lost forever. What do you suppose I'll have to do? Hey! What's the matter with my garage? That's what it will be. What do you know, Joe?

T. What do you know?

C. This is the littlest boat. I have to build so many docks around here.

T. So many docks you have to build.

C. Yes, and all these are brothers. This is the best one [picks up middle boat], because, look. He can carry more sand than the other ones.

T. He's the best one of all.

C. You know what? You know what I have to do? See this boat here? That's the littlest one, and this one is the biggest one, and see, those are brothers. These two are brothers and these two are brothers.

T. They're all brothers.

*T, Therapist. †C, Child.

C. Yes. And they all have some docks, but he [middle boat] has the coziest one.

T. He has the nicest, coziest one.

C. And this [big boat] and this [little boat] each have one, but he [middle boat] can carry nice soft dirt for people to the lake. We better not use this boat [tiny boat]. You know, this is the guy who stands and has to watch all of these things. [Picks up policeman figure and gestures toward boats]. And this one. You know what? I'm pretending that this is an oil place and that's where the boats get their power. You know, they don't have any power when they start off. They come around and after awhile they put their boats in this space where they can get power.

T. That's what they do.

C. I know what I'm doing. You know what I'm pretending? This is all the family, the whole family, the whole family. This is the family.

T. You're pretending it's the whole family.

C. Yeah. I have to. Well, what do you know? What do you know, Joe? What do you know, Joe? Linga, linga, linga, linga, linga, linga. Hey! I'm pretending. You see these cowboys? They're the guards.

T. They are the guards, huh?

C. All of them. They're the guards of these garages.

T. They guard them.

C. They guard them. There's the guard. You see, if anyone comes around to steal the boats . . . well, they'll shoot them.

T. They shoot anyone who tries to steal the boats.

C. There's another guard. He's a husky one. Look, there's sand in this boat [middle one]. He carries it well. Golly, he's stuck in a ditch, but he's cozy. Cozy, but him and him . . . no, really every three of them are the coziest.

T. Every three of them.

C. This is the best cowboy and he guards this boat [the middle boat]. Linga, linga, linga. You know what? They watch to see that no one steals anything. They watch the garages, too. One guard's in front of each garage.

T. One in front of each garage.

C. See, these are very lucky because they have guards. They're very lucky.

T. Very, very lucky because they have guards.

C. No one else has guards but them.

T. No one but them.

C. This guard watches this [big boat]. This guard watches this [middle boat]. And this guards this one [tiny boat]. This guy [middle boat] is lucky. He's lucky because he has the nicest house, the nicest house of all. He has the best house of all. He can just fit in right well. These [the other two boats] are lucky, too. They have power. They squeeze right in. He has power, too. He goes over and gets his power.

T. You have just a short while longer to play.

C. There he goes. See? He goes to his brother's place.

T. He's going right to his brother's place.

C. Hey, ring, ring! It's me. I'm your brother. It's all right. I was here before you, but come with me. Hey! Big boat and little boat say, "Please give me some power," and middle boat says, "O.K." Middle boat: "I'm going to get more power." Hey, Joe. Come on. I'll help you. We've got the best house in the world. We'll get some power. All we'll have to do is back right out and get it. We can get our power and gas easy. Linga, linga, linga. We can go now. When others come in, they'll be able to see that I built all this.

T. They'll see it right there where you built it.

After the third play session, which was very similar to the second, Tommy said he felt that he would not have to come back any more. He then went to his mother, she reported, and said: "Look here, Mother. There are some things that are mine, really mine. They belong just to me, and there are other things that I can share and will gladly share." His mother responded: "Of course, Tommy. That's the way it will always be."

The nursery school staff and Tommy's parents reported that Tommy had once again become the affable, free, expressive child they had known before.

THE CASE OF SUSAN

The nursery school staff described Susan, three years old, as a charming youngster whose winning smile and understanding ways had made her popular with both children and adults. Her mother had considered Susan's relations with her parents and an older sibling excellent. When Susan was three years three months, a new baby arrived in the family. Two days after the mother and the new baby daughter arrived home, Susan became babyish, immature, and whining at home. This behavior became evident in the nursery school as well. Susan's mother frantically telephoned the nursery school one day to ask whether something could be done to stop Susan's constant whining, which had annoyed everyone in the family. She could not understand how so wonderful and confident a child could have become a whimpering, clinging one in so short a time.

The nursery school staff referred Susan to the play therapist, who conducted three play

sessions with her. During the first two sessions Susan appeared to project her negative and hostile feelings for the new baby onto the humanlike balloon figure in the therapy room, throwing it on the floor, stepping on it, squeezing its head and face, and crushing it inside the vise. Once her feelings were recognized, accepted, and at least partly clarified, she proceeded in the last session to pick up the balloon figure, kiss it, toss it in the air, and dance around the room while she held it in her arms.

A transcript of recordings of the three sessions follows.

First Session

T. You can use these in any way you like. [Mother starts to leave room, and C looks at her.]

C. No, you stay here for awhile.

*M. Watch the watch. When this hand gets over here, I'll be back.

C. O.K. I'll bounce two balls.

T. Two at a time you did.

M. I'll lay down the watch where you can watch it.

C. O.K. . . . It's not ticking.

M. Want me to put it on you? [Places watch on C's wrist.] Bye.

T. You can just keep your eye on that watch.

C. [Waves good-bye to mother.] Where's the baby?

T. Where do you suppose the baby could be?

C. Here? That's the baby. Lookit the big baby. This is a balloon head. Mr. Balloon Head. [Picks up a balloon in the form of a human figure. Squeezes balloon and cries, "Mommy, Mommy, Mommy."

T. That's what it cries. Mommy, Mommy, Mommy.

C. [Continues to squeeze balloon and cry, "Mommy, Mommy, Mommoy." Looks at T and places balloon on table. Turns handle of vise.] What is it?

T. You want to know what it could be. It can be anything you want it to be.

C. A can opener.

T. Is that what it is? A can opener?

C. Lookit these soldiers. Are these cowboys or soldiers?

T. What do they look like?

C. Cowboys. See the cowboys. Those are all cowboys.

T. Mm-hm.

C. I'm gonna be a monkey.

T. That's what you're going to be.

C. [Puts on monkey mask.] Lookit me. I'm the monkey.

T. Susan is the monkey.

*M, mother.

C. Now I'm gonna be a piggy. I'll say oink-oink. [Puts on pig mask.] Oink, oink, oink, oink, oink.

T. The piggy goes oink, oink, oink.

C. Oink, oink, oink, oink. [Takes mask off.] Now I'm gonna be a clown. This goes oink, oink, oink, too. Oink, oink.

T. The clown goes oink, oink.

C. Oink, oink, oink. [Laughs.] Now I'm gonna be a baby and drink from the bottle of water. Shall I?

T. That's up to you.

C. Should I sprinkle here? Here. Open your hand.

T. You want to sprinkle in my hand.

C. [Sprinkles water in T's hand.] Rub them together. [Drinks from the bottle and then replaces it on bench. Turns handle of vise again.] Now I have to can opener this. [Puts figure balloon into vise; it squeaks.] She doesn't want to be can opened.

T. She doesn't?

C. No. I heard my Mummy walking. Hey! It's almost up to here. This number right here. [Indicates number on wrist watch.] I hear her coming.

T. You hear her coming.

C. [Turns handle of vise. Shakes it back and forth. Looks at nursing bottles. Again turns vise handle and looks out window. Picks up figure balloon and squeezes it; drops it and steps on it.] I'm gonna throw the ball. You kick it like that.

T. Mm-hm. That's what you do to it.

C. See what you do? Rocky-rocky the baby to sleep. Where's the baby? Where is she? Here's a mirror.

T. Mm-hm.

C. [Peers in mirror of dresser.] Tick-tick.

T. That's the way it goes.

C. Tick-tock. Tick-tock. There's the baby in there. [Points to doll house.] Baby walking upstairs. One, two, three, four, five. Into your beds. They're in their beds. Into your bed you go, bad girl. [Baby doll.] And this one is a big girl.

T. A big girl.

C. With a round head, Walk, walk, walk. Here's the daddy going to bed now. Walky, walky, walky. Right next to the girl. [Middle doll.]

T. Mm-hm.

C. And here's the mommy. Walky, walky, walky. Right next to the baby. [Undresses male doll.] I'm taking his pantsies off.

T. Mm-hm. You're taking his pants off.

C. Walk, walk, walk. Walking up to bed, walking up to bed. Three little children.

T. Three little children and two big people.

C. And another little baby. Here's me. I'm taking her clothes off. I'm going to bed. Now he's up. [Male doll.] Up and up and up. [Dressing male doll.] Little up, little up. Put your pants back on. Walky, walky, walky downstairs. [Female doll.] Walky, walky, walky downstairs. [Male doll.] Walky, walky, walky. [Another male doll.] [Walks baby doll downstairs.] I'm climbing up

this ladder. Let's climb up the ladders. Just climbing up the ladders.

T. Mm-hm.

C. And the little one on top of the bed. [Middle doll.] The little one sleeps under the bed. [Baby doll.]

T. One on top, one underneath.

C. Two underneath. Here's the bedroom. [Bends figure balloon and squeezes it.] I like that noise. Squeak, squeak, squeak. This is a big bed. Here's your bed.

T. That's my bed, huh?

C. Who sleeps in that bed?

T. Anyone you want.

C. Me. This is my little chair. [Crouches and sits on bed.] How come this doesn't go? [Points to watch on arm.]

T. You wonder why it doesn't go.

C. Oh. Supper is open. Here's your supper.

T. Quite a supper.

C. That's my mother's watch. Just pretend it's your supper. [Turns handle of vise.] Zoom, zoom, zoom, zoom. Here's your supper. Eat it up. Don't eat my mother's watch up. Just eat your supper up. Zoom, zoom, zoom. Here's your watch. Zoom, zoom, zoom. Here's your watch.

Second Session

C. [Talks to mother.] Are you gonna stay here? Here's a balloon. [Waves a ballon figure at T.] Good-by. [To mother.] Mommy, leave your watch here. I want to see what time it is. [Looks at T's watch.] It will still be there tomorrow. [Puts balloon figure in upper part of doll house. Empties bag of dolls.] There. In the garbage can. [Walks to nursing bottles.] I'm gonna drink from this. [Drinks from large bottle and replaces it on bench.] He's gonna shoot you. [Cowboy figure.] Bang. Hoppy is gonna shoot you. Shoot and tie you up. All cowboys are shooting. [Handles soldiers and shoots T a few times.] Everyone is shot. [Squeezes figure balloon and it squeaks. Walks figure balloon up the stairs of doll house.] Walk, walk, walk. [Throws figure balloon aside. Sits on floor, fingers stairway. Picks up figure balloon and whispers.] She's going to sleep. Shall I take her head off?

T. That's up to you.

C. [Places balloon in box with blocks.] That's a block. [Picks up figure balloon again, brushes it against T's face.] I wanta take your glasses off.

T. You'd like to do that, but that's one thing you can't do here.

C. Let's pretend to play school. O.K.? And you're the teacher. O.K.?

T. And I'm the teacher.

C. [Cuts a piece of paper and folds it in half. Cuts paper along folded line and into quarters, looks at T, folds paper again and shows it to T.] I'm gonna give these to my mother. That's for Mother's Day.

T. Is that a Mother's Day present?

C. Yeah. [Cuts another piece of paper in two.] This is my mother's present, too.

T. You have quite a few to give your mother. You like to give her things, huh?

C. [Holds papers in hand.] These are my mother's and my daddy's, too. Just for my mother and daddy. [Lifts come-back toy.] He's a big clown.

T. Mm-hm.

C. [Carries come-back toy to T.] There. Walk, walk, walk, walk, walk.

T. There you go.

C. [Leans against come-back toy, pushing it down.]

T. You want it to go down.

C. Yeah. [Pushes come-back toy into sandbox.] He's crying.

T. You're making him cry.

C. Yeah. [Finally succeeds in pushing come-back toy into sandbox.]

T. There, you have it.

C. He's crying. Nobody's taking him out.

T. He is just going to stay in there all the time.

C. [Hands T papers that she has cut.] Will you fold these for my mother? And my dad.

T. For no one else.

C. Not even you.

T. Not even me.

C. No. [Starts to cut paper again. Continues; folds one half sheet in two again and places it on top of the others.] See? Some's for my family, and not for you, either.

T. Not for me.

C. No. [Cuts more paper.] Only one is for you, and this is all you're getting. Here. None for your mother. No. It's all mine and my mother's.

T. Just yours and your mother's.

C. I'll be the teacher and gather up your things. O.K., honey. Let's, honey. Yes, honey. Where's that paper? And you, honey. Honey, honey. I'm gonna sprinkle some. [Drops a handful of sand from sandbox into pail of water. Watches it. Takes more sand and drops it into pail. Looks at T and laughs.] It's all getting brown, isn't it? [Throws more sand into pail.] The floor is getting wet. [Continues to drop handfuls of sand into pail.] It's getting brown water.

T. Yes, it is. It's getting to be brown water.

C. [Throws more sand into pail.] I splashed my shoe. See? [Sprinkles some sand over come-back toy. Drops more sand into pail and waves her hands in the air.] I wanta go wash them.

T. You want to wash them? O.K.

C. [Leaves room with T.]

Third Session

C. [Waves good-by to mother and runs into the room. Drops a handful of sand into pail of water.]

T. It went right in, didn't it?

C. Look how much. [Drops another, larger handful into pail and laughs.] A big splash. Splash. The water's getting brown. [Drops two more handfuls of sand into pail.]

T. It's getting browner and browner.

C. Mmm. Now I'm making pie. [Plays in sand box.]
T. So you're making a pie, that's what.
C. Here's your pie.
T. Is that for me?
C. Mm. Take a shovel and eat it. Take a spoon and eat it.
T. You want me to eat it.
C. [Thows more sand into pail and smiles at T.] O.K. Here. [Gestures toward T with shovel.]
T. You want me to eat with that, huh?
C. Not really.
T. You just want me to pretend?
C. Yes.
T. Mm-hm.
C. I'm throwing it all over the floor. [Throws sand on the floor.]
T. You are?
C. Swoosh, swoosh, swoosh.
T. You really like that.
C. [Continues to play with sand.] Now this is a little cookie, and I'll put it in a plate.
T. Mm-hm.
C. Here. And I'll give you some more. [Throws more sand into pail.] It's getting dark blue.
T. That's what it's getting to be. Dark blue.
C. [Fills mold with sand, pats it, and gives it to T.] Eat it.
T. You want me to eat it now.
C. And then I'll give you some more. Eat it up. O.K. Now eat it.
T. Now you want me to eat it. Suppose that I don't want any more?
C. Then you won't get any dessert. Now eat it all up. O.K., now take it, Now pick it up now. Hello, hello, hello, hello. [Dials telephone.] Pretend I hear the phone bell ringing and I say hello, and you talk.
T. Oh, all right, we'll pretend that.
C. Hello?
T. Hello.
C. Who is this?
T. Who is that?
C. This is Susan, and she's playing here. Good-bye.
T. Good-bye.
T. [Sneezes.]
C. God bless you.
T. Thank you.
C. [Picks up large bottle and drinks. Replaces it on bench.] I like to play here.
T. You like coming here and playing.
C. [Walks over to balloon figure and kisses it. Tosses it into air and catches it several times while dancing around the room.]
T. Well, our time is up for today, Susan.
C. One more bouncy and I'll go up.
T. O.K., one more bouncy and we'll go.
C. [Throws balloon figure into air one more time. Lets it fall on floor.] O.K. Goodbye, good-bye. Good-bye, Mister.
T. Good-bye.

Susan's mother came in after this last session to tell the therapist that she was very happy about Susan, who had again become a pleasant child; that she was no longer afraid to leave Susan with the baby; that Susan showed more affection for the baby and had assumed some responsibilities in the baby's care.

BENEFITS OF PLAY THERAPY FOR THESE TWO CHILDREN

Both Tommy and Susan used symbolic forms—boats and balloon figures—to localize their anxiety about a new baby in the family. The gains in both cases were growth in terms of emotional insight, a feeling of security and comfort within themselves. What they needed was an opportunity to express their attitudes in an accepting relationship where they could feel respected as individuals, whatever their feelings and perceptions about themselves and their families might be.

Situational play therapy provided these children with an opportunity to work out temporarily disturbing feelings, and so removed the possibility that these feelings would be repressed, lose their identification to reality, and perhaps eventually seriously damage the self by pervading it with free-floating or systematic anxiety. Freed from these temporarily disturbing feelings, these children were able to use their psychic energies to express their real selves, thereby freeing these energies for a fuller utilization of their potentialities in both personal and social situations.

Received October 9, 1950.

IQ VARIABILITY IN RELATION TO AGE ON THE REVISED STANFORD-BINET

MURRAY ABORN
MICHIGAN STATE COLLEGE
AND
GORDON F. DERNER
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

THE IQ is generally regarded as a relatively stable index of intelligence. Its stability, however, is dependent upon the degree to which means and standard deviations remain constant from one age level to the next. For example, if the standard deviation of IQ's is 12.5 at year 6 and 20.0 at year 12, even if the mean is a constant of 100 for both ages, an IQ of 112.5 at year 6 represents the same degree of brightness as an IQ of 120 at year 12. When the means and standard deviations vary from age to age, therefore, there can be no uniform interpretation of the IQ.

In the standardization group of the 1937 Revision of the Stanford-Binet, the fluctuations of mean IQ amoung the various age levels range from 100.9 at year XIV to 109.9 at year II-6 [10, p. 35], and the fluctuations of standard deviations among the various age levels range from 12.5 at year VI to 20.6 at year II-6 [10, p. 40]. Many of the differences between means and standard deviations involved in the age-to-age fluctuation are statistically significant [1], and the IQ values are thus not directly comparable. In the absence of positive proof to the contrary at the time, Terman and Merrill [10, pp. 38-41] proposed that the unequivalences were due to sampling errors,[1] the implication being that with adequately sampled data, the means and standard deviations for all age levels would be the same. Along with McNemar, however, they also suggested that artifacts inherent in the structure of the test may be responsible. McNemar's position [8, p. 85] is rather fully explained in terms of the curves for percentages passing the various subtests at particular age levels, and in terms of levels of difficulty.

Of the two possible explanations, added evidence for the artifact hypothesis is contained in a study by Goodenough [6]. She reported two experiments in which the variability among standard deviations at different age levels was comparable to that found in the Terman and Merrill standardization data. No tests of significance between the two sets of data were made, but the similarity was nonetheless evident. Goodenough also reported her data to be in concurrence with that of Ebert [4], and like McNemar, she concluded that the difference between standard deviations derived from different age groups on the Revised Stanford-Binet are the result of artifacts within the testing instrument.

PROCEDURE

The present study was designed to test whether differences in IQ variability from one age level to another can be accounted for in terms of the sampling employed in the standardization of the Revised Standford-Binet, or are an artifact of the instrument. The original standardization data are compared to a new and extensive sampling. Further, the comparison between the new and the original data is subjected to a more thorough statistical analysis than that reported in previous experiments.

The specific questions raised are:

[1]The question of the value of norms based upon a faulty sample is not raised by the test authors.

1. Are there significant differences in IQ variability age by age between a new sample and the original standardization sample?

2. Are the significances of differences in IQ variability between the age levels of the new sample comparable to the significances of differences in IQ variability demonstrated for the age levels of the original standardization data?

Over a period of several years, the authors developed a file of some six thousand Revised Stanford-Binet, Form L, records on subjects from metropolitan New York. These examinations were administered by advanced graduate students in clinical psychology. Their work was closely supervised, and every record booklet was re-examined by a qualified psychologist. In the students' training an emphasis was placed on a strict adherence to administrative and scoring procedure [3].

For purposes of this study, a selection of cases from the file was made. Cases were selected according to the following criteria:

1. A chronological age between 4-5 and 14-2. This is the age of greatest usefulness for the Revised Stanford-Binet.

2. For subjects below six, only those who were within one month of their birthday or half-year birthday were selected. The subjects selected above six were within two months of their birthday. This procedure closely approximates the standardization procedure [8, p. 7].

3. All subjects were white and American-born, as in the standardization population.

4. A selection of cases which most nearly approximated the occupational groupings of the standardization population.

TABLE 1

PERCENTAGE DISTRIBUTION OF OCCUPATIONS OF FATHERS OF THE SUBJECTS IN TERMAN AND MERRILL'S STANDARDIZATION POPULATION AND IN THE POPULATION OF PRESENT STUDY

Occupational Group*	Terman and Merrill ($N = 2757$)	Present Study ($N = 784$)
I	4.5	17.7
II	7.8	16.1
III	25.5	17.8
IV	14.9	0.6
V	31.4	33.9
VI	9.4	7.4
VII	6.6	6.5

*According to the Goodenough classification [5], in which Group I is professional; Group II, semiprofessional and managerial; Group III, clerical, skilled trades, and retail business; Group IV, farmers; Group V, semiskilled occupations, minor clerical positions, and minor businesses; Group VI, slightly skilled trades; and Group VII, day laborers.

RESULTS

Table 1 gives a comparison of the occupational groupings of the fathers of the subjects used in the standardization sample and in the present stndy according to the Goodenough classifications [5, pp. 510-512] used by Terman and Merrill. Since the subjects of the present study were drawn from metropolitan New York, no rural group is included. It will also be noted that the present study contains a considerably greater number of professional and semi-professional fathers and a somewhat smaller number of fathers in skilled trades and in retail business than the standardization sample. The fact that the present sample is weighted in the direction of the higher socioeconomic levels accounts for the uniformly higher mean IQ's at all the ages studied.

Table 2 presents the means, standard deviations, and coefficients of relative variation for the Terman and Merrill data and for the data of the present study. Because the mean IQ's for the present study are higher at every age level, T-tests of significance for differences between coefficients of relative variation [9, p. 418] were used in preference to tests of significance for differences between the standard deviations. No significant differences in variability between the Terman-Merrill and the present data were found at any age level. The highest T obtained was 1.10, which is significant only at the .14 level. Other T's ranged downward from .87, indicating even less significance. These results demonstrate that the same differences between various age levels occur in the case of two independently drawn samples, and the thesis that age-to-age variability is due to sampling errors is there-

TABLE 2

MEANS, STANDARD DEVIATIONS, COEFFICIENTS OF RELATIVE VARIATION, AND *T*-RATIOS FOR TERMAN AND MERRILL'S STANDARDIZATION POPULATION AND THE POPULATION OF THE PRESENT STUDY

Age	Number Cases		Mean		Standard Deviation		Coefficient of Relative Variation		
	Terman and Merrill	Aborn and Derner	Terman and Merrill	Aborn and Derner	Terman and Merrill	Aborn and Derner	Terman and Merrill	Aborn and Derner	*T*-Ratio*
4½	79	23	103.7	125.1	16.2	17.9	15.6	14.3	0.52
5	90	24	104.3	114.6	14.2	14.4	13.6	12.6	0.47
5½	110	37	101.4	113.5	14.3	13.9	14.1	12.2	1.10
6	203	54	101.0	114.7	12.5	13.0	12.4	11.3	0.87
7	202	100	101.8	115.6	16.2	17.7	15.9	15.3	0.44
8	203	100	102.4	114.6	15.8	16.8	15.4	14.7	0.53
9	204	83	104.3	110.7	16.4	18.5	15.7	16.7	0.64
10	201	100	104.5	113.7	16.5	19.4	15.8	17.1	0.87
11	204	100	104.3	111.9	18.0	18.6	17.3	16.6	0.47
12	202	87	103.6	109.6	20.0	19.6	19.3	17.9	0.82
13	204	42	104.0	113.1	17.9	18.5	17.2	16.4	0.39
14	202	34	100.9	115.1	16.1	18.0	16.0	15.6	0.19

*Between coefficients of relative variation.

TABLE 3

TESTS OF SIGNIFICANCE BETWEEN COEFFICIENTS OF RELATIVE VARIATION FOR AGE LEVELS OF THE REVISED STANFORD-BINET FOR THE ORIGINAL STANDARDIZATION DATA OF TERMAN AND MERRILL AND THE DATA OF THE PRESENT STUDY†

Chronological Age	Chronological Age										
	5	5½	6	7	8	9	10	11	12	13	14
4½	0.60 (1.23)	0.81 (0.95)	1.25 (2.29)**	0.42 (0.20)	0.17 (0.14)	0.95 (0.07)	1.12 (0.07)	0.93 (1.10)	1.40 (2.83)**	0.75 (1.05)	0.45 (0.45)
5		0.17 (0.36)	0.61 (1.00)	1.26 (1.76)*	0.99 (1.41)	1.80 (1.62)*	2.03* (1.68)*	1.82 (2.72)**	2.29* (3.99)**	1.46 (2.65)**	1.12 (1.82)*
5½			0.50 (1.48)	1.71* (1.43)	1.40 (1.05)	2.30* (1.27)	2.58** (1.35)	2.35** (1.35)	2.85** (3.74)**	1.81* (2.37)**	1.42 (1.50)
6				2.56** (3.43)**	2.22** (3.03)**	3.12** (3.27)**	3.49** (3.33)**	3.25** (4.50)**	3.71** (5.90)**	2.38** (4.44)**	1.94 (3.50)**
7					0.39 (0.45)	0.81 (0.18)	1.08 (0.09)	0.80 (1.17)	1.46 (2.66)**	0.51 (1.03)	0.14 (0.09)
8						1.17 (0.27)	1.46 (0.36)	1.18 (1.61)*	1.83* (3.12)**	0.81 (1.54)	0.41 (0.53)
9							0.22 (0.09)	0.06 (1.33)	0.62 (2.81)**	0.13 (1.26)	0.47 (0.26)
10								0.29 (1.25)	0.48 (2.73)**	0.32 (1.17)	0.65 (0.17)
11									0.70 (1.50)	0.09 (0.08)	0.44 (1.07)
12										0.65 (1.59)	0.96 (2.58)**
13											0.30 (1.00)

†In each cell, the upper figure represents data from the present study. The lower figure, in parentheses, represents the Terman and Merrill data for comparison. Significance determined by t-distribution where N is less than 30, and by T-distribution where N is greater than 30.

*Significant at the .05 level.

**Significant at the .01 level.

fore contraindicated. Variability must be due to some internal factor in the test.

In a further analysis of the differences in variability between age levels, intertests of significance between coefficients of relative variation for the ages levels of the standardization sample and for the present study were made. The *T*-ratios are presented in Table 3. Twenty-six of the comparisons between the age groups of Terman and Merrill's data show significant differences. Seventeen of the comparisons between the age groups of the present study's data show significant differences. Of the 66 age comparisons in all, the data of Terman and Merrill and of the present study are in agreement on significance or nonsignificance in 51 instances. The results of this inter-age level analysis again indicate that the differences in variability cannot be accounted for solely on the basis of sampling inadequacies. The differences are the result of features inherent in the test construction.

CONCLUSIONS

The demonstrated differences in variability between the age levels of the Revised Stanford-Binet as an artifact of the test affects the constancy of the IQ and requires that some correction for age differences be made. McNemar presents an adjustment table for atypical variability [8, pp. 172-174]. Only those age groups with the most divergent variability are listed. Goodenough [6] suggests the use of standard scores based upon an assumed mean of 100 and an assumed standard deviation of 17. Although the use of standard scores is in keeping with modern psychometric opinion [2, 7, 11], Goodenough's assumed constants give scores which are too apt to be confused with the IQ's themselves. *T*-scores (sometimes called *Z*-scores), with an arbitrary mean of 50 and an arbitrary standard deviation of 10, would seem to be most suitable. *T*-scores can be readily computed from the data provided by Terman and Merrill [10, p. 35, p. 40]. Tables of converted scores have been prepared by Derner and Aborn [3]. Terman and Merrill have prepared a table of standard scores [10, p. 42], but they ignored the very fluctuations in variability which make standard scores essential.

SUMMARY

In order to investigate whether the differences in variability between the age levels of the Revised Stanford-Binet are due to sampling errors involved in the original standardization data or to artifacts in the test structure, a sample of 784 cases between the ages of 4½ and 14 was selected from a file of some six thousand cases tested in metropolitan New York. The socioeconomic level of these subjects was above that represented in the standardization sample, resulting in a higher mean IQ for the sample of the present study.

Tests of significance between coefficients of relative variation for the standardization data and for the data of the present study revealed no significant differences between the two. Intercomparisons between the age levels of the standardization sample and between the age levels of the present study sample show high agreement.

The results of this study do not support the thesis that the age differences in variability are a result of sampling inadequacies. The results are in agreement with the position that these differences occur from features inherent in the structure of the test. In order to maintain the meaning of the IQ as a uniform index of intelligence, therefore, standard scores should be employed in preference to IQ's themselves. *T*-scores (sometimes called *Z*-scores) are the most useful form of standard scores for this purpose. Tables for converting Revised Stanford-Binet IQ's into *T*-scores are available [3].

Received August 17, 1950.

REFERENCES

1. BROWN, F. The significance of the IQ variability in relation to age on the Revised Stanford-Binet scale. *J. genet. Psychol.*, 1943, **63**, 177-181.
2. CRONBACH, L. J. *Essentials of psychological testing.* New York: Harper, 1949.
3. DERNER, G. F., AND ABORN, M. *The clinical diagnosis of intelligence.* New York: Authors, 1949. (Mimeo.)
4. EBERT, E. H. A comparison of the original and Revised Stanford-Binet scales. *J. Psychol.*, 1941, **11**, 47-61.
5. GOODENOUGH, FLORENCE L., AND ANDERSON, J. E. *Experimental child study.* New York: Century, 1931.

6. GOODENOUGH, FLORENCE L. Studies of the 1937 revision of the Stanford-Binet scale: I. Variability of the IQ at successive age-levels. *J. educ. Psychol.*, 1942, **33**, 241-251.
7. GOODENOUGH, FLORENCE L. *Mental testing.* New York: Rinehart, 1949.
8. MCNEMAR, Q. *The revision of the Stanford-Binet scale.* Boston: Houghton Mifflin, 1942.
9. PEATMAN, J. G. *Descriptive and sampling statistics.* New York: Harper, 1947.
10. TERMAN, L. M., AND MERRILL, MAUD A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
11. WOLFLE, D. Standard scores and percentile scores. *Amer. Psychologist*, 1947, **2**, 76.

THE STANFORD-BINET AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

ARDEN N. FRANDSEN
AND
JAY B. HIGGINSON
UTAH STATE AGRICULTURAL COLLEGE

FOR the psychological study of children in schools and in child guidance clinics, the most popular individual test has been the Stanford-Binet intelligence scale [2, p. 50]. This is a highly useful test because, besides providing reliable mental ages or intelligence quotients indicative of potentiality for successful school learning and problem solving, it can be made to yield in the hands of a skillful clinician important clues to differential abilities, work methods, and personal and social adjustment. There are available, of course, alternative tests to the Stanford-Binet, such as the Kuhlmann Tests of Mental Development, the Herring Binet Revision, and the Detroit Tests of Learning Aptitude, but they are not used nearly as frequently [2, p. 50]. Because of the great popularity [2, p. 50] of the Wechsler-Bellevue Intelligence Scales, the new Wechsler Intelligence Scale for Children [3], which "has grown logically out of" the former, may become a more important competitor to the Stanford-Binet when an individual intelligence test is needed for the ages from five through fifteen years. Like the W-B, the WISC includes a *Verbal Scale* comprising Information, Comprehension, Arithmetic, Similarities, Vocabulary, and Digit Span (an alternate test); and a *Performance Scale* into which are grouped Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes (with either Coding or Mazes as an alternate test). Verbal IQ's, Performance IQ's, and Full-Scale IQ's are derived "by comparing each subject's test performance not with a composite age group but exclusively with the scores earned by individuals in a single age group" [5, p. 3].

A comprehensive evaluation of the WISC will require, besides the standardization data already reported by Seashore and others [3], studies of the test's validity for predicting school achievement at different levels, its possibilities for indicating differential aptitudes, quality of work methods, and symptoms of personal and social adjustment. As a contribution to such an evaluation, this investigation will report data from children at the fourth-grade level on (1) the correlations between Stanford-Binet and WISC IQ's, (2) on the comparability of their IQ norms, and (3) their validity for predicting school achievement as measured by the Stanford Achievement Test, Intermediate Battery, Form G.

PROCEDURE

During the 1949-50 school year the junior author administered both the Revised S-B and the WISC to 54 unselected fourth-grade children (two sections constituting the entire grade in the school), ranging in age from 9-1 to 10-3, having average ability (Table 1), and approximating the average in achieve-

TABLE 1

INTERCORRELATIONS, MEANS, AND SIGMA'S OF S-B AND WISC IQ'S FOR 54 CHILDREN

	S-B	W-F	W-V	W-P
WISC, Full Scale	.80			
WISC, Verbal Scale	.71	.94		
WISC, Performance Scale	.63	.88	.52	
Mean	105.8	102.4	100.9	103.5
Sigma	11.15	11.15	12.25	11.20

TABLE 2
CORRELATIONS OF S-B AND WISC IQ'S WITH EDUCATIONAL AGE ON THE STANFORD ACHIEVEMENT TEST FOR 54 CHILDREN

	Read.	Lang.	Arith.	Lit.	Soc. St.	Sci.	Spell.	Total
WISC-F	.68	.55	.64	.75	.71	.45	.55	.76
WISC-V	.55	.49	.56	.59	.52	.60	.48	.62
WISC-P	.62	.41	.61	.58	.64	.53	.43	.65
S-B	.67	.64	.48	.47	.61	.46	.45	.63
Mean	4.38	4.61	4.40	4.90	4.49	4.54	4.42	4.56
Sigma	0.95	0.98	0.57	1.32	1.60	1.00	1.00	0.77

ment (Table 2). To control practice effects, the two tests were administered alternately so that one child was given the S-B as the first test, and the next child the WISC as the first test. Both tests were administered within ten days. The examiner, a graduate student in psychology, administered the tests after an intensive course in individual intelligence testing, following explicitly the directions for the tests as given in their respective manuals. The two teachers of the two classes of fourth-grade children, both of whom had had previous experience with group tests, administered the Stanford Achievement Test to their pupils at the end of the school year, and the tests were scored to yield total educational ages and separate educational ages for each subtest in the test battery.

RESULTS

The intercorrelations between the S-B IQ and the WISC IQ's, and the means and standard deviations of the IQ distributions for the 54 children are presented in Table 1.

As Table 1 indicates, the WISC Full Scale and the S-B are measuring to a considerable extent the same factor or factors; the coefficient of determination ($r^2 \times 100$) for the correlation of .80 indicates that 64 per cent of the variance in the S-B IQ distribution is associated with the variance in the WISC IQ distribution. For the W-B adult scale, correlations with the S-B ranging from .62 to .93 have been reported [4, p. 134]. Between the WISC Verbal and Performance scales, for our subjects, the independence or extent to which they measure different factors is marked. The coefficient of determination for the correlation of .52 indicates that only 27 per cent of the variance on the one scale is accounted for by the variance on the other. For Wechsler's standardization subjects of comparable age [5, p. 11], however, the verbal-performance correlation was found to be higher ($r = .68$). Both scales, however, correlate highly with the Full Scale, of which, of course, they are parts. And both the WISC Verbal and Performance IQ's correlate appreciably with the S-B, the largely verbal S-B IQ agreeing more closely with the WISC Verbal IQ ($r = .71$) than with the WISC Performance IQ ($r = .63$).

Also from Table 1, it is apparent that our sample of subjects is about average in general mental ability, though slightly more homogeneous than both Terman's and Wechsler's standardization populations, for which the IQ standard deviations were 16 and 15, respectively. The mean IQ of 102.4 on the WISC approximates very closely the mean IQ of 105.8 on the S-B; the t ratio of 0.45 for this difference of 3.4 IQ points indicates that it is not significant statistically.

Data for comparing the validities of the S-B IQ and the WISC IQ's for predicting the school achievement of our fourth-grade subjects, as measured by the Stanford Achievement Test, are presented in Table 2.

Recognizing that intelligence is only one factor affecting school achievement, the validity ($r = .63$) of the S-B in prediciting the total achievement scores of our subjects on the Stanford Achievement Test is relatively high, compared to the intelligence-achievement correlations typically reported [1, p. 846]. But in this study the validity of the WISC Full Scale IQ is even higher: $r = .76$. The correlations with total achievement of both the WISC Verbal Scale and the WISC Performance Scale approximate very closely the cor-

responding correlation with the S-B. For the subtests of the Stanford Achievement, the correlations are also generally slightly higher with the WISC. Only in the instance of Language Usage, for this population, is achievement predicted better by the S-B.

SUMMARY AND CONCLUSIONS

The new Revised Stanford-Binet Tests of Intelligence, Form L, the Wechsler Intelligence Scale for Children, and the Stanford Achievement Test, Intermediate Battery, Form G, were administered to a sample of 54 fourth-grade children (the entire fourth grade of one school). An evaluation and comparison of the mean IQ's from the intelligence tests and of the intercorrelations both between the intelligence tests and with achievement lead to the following conclusions:

1. IQ norms from the S-B and WISC are comparable at the average level, and very probably fairly comparable at least within the range of one to two sigma's above and below the mean.

2. The correlation of .80 between the S-B and the WISC Full Scale IQ's indicates that to a considerable extent the two tests are measuring the same factor or factors. Of the variance in the S-B IQ distribution, 64 per cent is associated with the variance in the WISC IQ distribution.

3. The factors measured by the Verbal and Performance scales of the WISC are, for our population, fairly independent ($r = .52$). The Stanford-Binet correlates more closely with the WISC Verbal Scale ($r = .71$) than with the WISC Performance Scale ($r = .63$).

4. Recognizing that intelligence is only one factor affecting school achievement, both the S-B and the WISC Full Scale (and the subscales of the WISC) are valid as predictors of achievement as measured by the Stanford Achievement Test. For this population the validity was found to be higher for the WISC ($r = .76$) than for the S-B ($r = .63$).

Received September 11, 1950.

REFERENCES

1. EURICH, A. C., AND CAIN, L. F. Prognosis. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* New York: Macmillan, 1940. Pp. 838-860.
2. LOUTTIT, C. M., AND BROWNE, C. G. The use of psychometric instruments in psychological clinics. *J. consult. Psychol.*, 1947, **11**, 49-54.
3. SEASHORE, H., ALEXANDER, W., AND DOPPELT, J. The standardization of the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1950, **14**, 99-110.
4. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
5. WECHSLER, D. *Wechsler Intelligence Scale for Children.* New York: Psychological Corp., 1949.

THE PERFORMANCE OF CONGENITAL SYPHILITICS ON THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN[1]

FLORENE M. YOUNG AND VIRGINIA A. PITTS
UNIVERSITY OF GEORGIA

NUMEROUS writers have listed congenital syphilis as a cause of mental deficiency [2, 3, 4, 8, 9]. Some physicians dealing with congenital syphilitics now question the accuracy of this viewpoint. Actually, there is a dearth of scientific research on the problem, and investigations are needed.

PROBLEM AND PROCEDURE

The present study reports the responses of 40 juvenile congenital syphilitics and 40 control subjects, as obtained from administering the Wechsler Intelligence Scale for Children. The 40 experimental subjects were diagnosed as congenital syphilitics by the staff of the Alto Medical Center, Alto, Georgia. The children were in the hospital for treatment at the time of testing. The sample consisted of 21 Negro girls and 19 Negro boys.

Since the WISC was standardized on white subjects, it was considered advisable to obtain a control group. The members of the latter were paired with the syphilitic subjects on the basis of age, sex, race, geographic region, and paternal occupation. The control subjects attended a rural school for Negro children in a country located in northeast Georgia. The matching for age was within six months.

Subjects ranged in age from 6 to 16 years. The mean CA of the experimental group was 12.02 years, and the standard deviation, 2.32 years. The mean CA of the control group was 12.09 years, and the standard deviation, 2.36 years. All subjects were rural children, and the paternal occupations fell within categories XI, XII, and XIII of the Census, that is, farm laborers and foremen, and laborers. Wechsler lists these workers as Occupational Group 8 [11, p. 9].

All tests were given the experimental subjects by the junior author in a private room at Alto Medical Center. Special care was taken not to test any child who had a headache or who expressed any complaints or objections. The same examiner tested the control group, provision having been made at the school for adequate examination facilities.

The subtests employed are indicated in Table 2. Since six tests of the Verbal Scale were used, prorating was necessary, as indicated in Wechsler's manual [11, p. 23].

RESULTS

The congenital syphilitics have records in-

[1]This report is based upon a thesis prepared by the junior author and submitted to the Graduate School, University of Georgia, June 16, 1950.

The authors thank Dr. E. M. Christensen, Medical Director, Alto Medical Center, Georgia, for his cooperation and advice in connection with the testing of the experimental cases.

TABLE 1
INTELLIGENCE QUOTIENTS FOR FORTY CONGENITAL SYPHILITICS AND FORTY CONTROLS ON THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

	Syphilitics				Controls				
	Mean	Median	σ	σ_M	Mean	Median	σ	σ_M	CR
Performance	59.00	58.25	12.40	1.96	67.38	68.25	15.15	2.39	2.71
Verbal	67.75	65.10	10.60	1.67	79.00	78.25	14.50	2.29	3.98
Full Scale	59.00	57.80	12.40	1.96	69.80	71.04	15.08	2.38	3.51

ferior to the control subjects. Their IQ's on the Performance Scale, Verbal Scale, and Full Scale are 8 to 11 points lower than those of the control group (Table 1). The differences between the groups on these scales are significant at the 1 per cent level.

According to Wechsler's classification, the syphilitic group, with a mean IQ of 59 is "mentally defective," since he regards an IQ below 70 as evidence of feeblemindedness [11, p. 16]. The control group, with a mean IQ of 69.8, is "borderline," by Wechsler's classification [11, p. 16].

The control group, however, was selected as a representative sample of rural Negro children in the state. It seems unreasonable that these subjects should be regarded as "borderline." Certainly they would not be so classified by socioeconomic criteria, nor by the judgments of observers. According to the standards of their culture these children seemed to be a typical group.

Inasmuch as Wechsler's test was standardized only on white children, it is apparent that the test norms are rather inapplicable to a sample of rural southern Negro children. An example of a possible handicap to such subjects is the fact that all human figures used for the Picture Arrangement and Object Assembly tests were those of white persons.

In consideration of these facts, the emphasis in the present study will be placed upon the relative achievements of the control and syphilitic groups, rather than upon a comparison of these subjects with Wechsler's norms.

From the obtained data it seems possible that syphilis may have been an important factor contributing to the significantly lower mental level of the experimental subjects. Nevertheless it should be emphasized that more research is needed. Nor should the possibility of other causative factors be disregarded.

For both groups, Verbal IQ's are higher than Performance IQ's. The differences are 9 and 12 IQ points for the syphilitic and control subjects, respectively, and the differences are significant at the 1 per cent level.

Seashore, who analyzed the records of Wechsler's subjects, has commented upon the difference between the results on the two scales: "One can expect an average discrepancy of about eight points between a subject's Verbal IQ and his Performance IQ on the WISC" [6, p. 319]. Seashore also found that Performance IQ's for rural subjects usually were higher than Verbal IQ's. Although the present study is based upon rural children, Verbal IQ's are significantly superior.

Since syphilitic subjects have markedly lower IQ's than do control cases it is appropriate to compare the subtest records of the two groups, thus ascertaining which tests best discriminate between them. The control group had better scores on all eleven subtests, and all but three comparisons resulted in differences which were significant at the 5 per cent level or better. The greatest differences between the groups occur in relation to the

TABLE 2

SCALED SCORES ON SUBTESTS FOR FORTY CONGENITAL SYPHILITICS AND FORTY CONTROLS ON THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

	Syphilitics				Controls				
	Mean	Median	σ	σ_M	Mean	Median	σ	σ_M	CR
Information	4.83	4.36	1.95	.31	7.05	6.21	3.30	.52	3.64
Comprehension	4.82	4.50	2.35	.37	5.87	6.00	2.81	.44	1.84
Arithmetic	4.75	4.42	2.14	.34	5.95	5.50	2.79	.44	2.14
Similarities	4.75	4.65	1.93	.30	6.48	6.25	2.71	.43	3.26
Vocabulary	2.17	2.00	2.04	.32	3.95	4.21	2.38	.38	3.56
Digit Span	6.37	6.70	2.97	.47	7.33	7.64	3.17	.50	1.39
Picture Comprehension	4.45	4.13	2.57	.41	5.90	6.50	3.06	.48	2.30
Picture Arrangement	3.46	2.50	2.97	.47	5.10	5.06	3.01	.48	2.45
Block Design	4.25	4.27	2.39	.38	5.13	5.36	2.53	.40	1.60
Object Assembly	4.42	4.70	2.30	.36	5.98	5.70	2.82	.45	2.74
Coding	3.40	3.27	2.32	.37	5.43	4.83	2.97	.47	3.38

following tests: Information, Vocabulary, Coding, Similarities, and Object Assembly. Comparisons indicate differences which are significant at the 1 per cent level (Table 2).

The first two tests, Information and Vocabulary, purport to measure ability to learn, but it must be conceded that they also are affected by the subject's cultural advantages. It is possible that the experimental cases had inferior environmental opportunities, although the subjects were paired on the basis of paternal occupation.

Cultural factors should exercise a minimal influence upon a test of immediate learning, such as the Coding test. The results show the syphilitic group to be significantly inferior in their learning ability, that is, in associating certain symbols with certain other symbols (CR = 3.38).

Landis and Bolles [5, p. 411] have pointed out that "The ability to generalize is one which matures slowly in the child." This ability is involved in the Similarities test on which the experimental group had relatively poor records (CR = 3.26).

The two groups also differed significantly on the Object Assembly test, the syphilitic subjects showing inferiority in dealing with whole-part relationships.

The three tests on which the two groups are most alike are: Comprehension, Block Design, and Digit Span (Table 2). CR's resulting from comparisons of group means on these tests do not reach the 5 per cent level of significance.

For both groups, the best records occur on the Digit Span and Information tests, while the poorest scores are to be found on the Vocabulary test.

Babcock [1] used vocabulary as a stable function in the investigation of intellectual impairment. Shipley [7] compared vocabulary and the ability to grasp the abstract as a measure of impairment. For the present study the Vocabulary test does not appear to be representative of the most stable function, since the two groups differ significantly with respect to this test. A relatively stable function would be one in which the groups do not differ significantly. Therefore, the Comprehension test has been employed in determining certain differences between subtests within the syphilitic group and within the control group, as shown in Table 3.

TABLE 3

CRITICAL RATIOS BETWEEN INDIVIDUAL SUBTESTS AND THE PERFORMANCE AND VERBAL SCALES ON THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

Test Comparison	Syphilitics CR	Controls CR
Performance–Verbal	3.40	3.51
Vocab.–Digit Span	7.38	5.45
Vocab.–Coding	2.51	2.43
Vocab.–Pict. Comp.	4.38	3.20
Compreh.–Inform.	0.02	1.74
Compreh.–Arith.	0.14	0.13
Compreh.–Simil.	0.15	0.98
Compreh.–Vocab.	5.41	3.31
Compreh.–Digit Span	2.58	2.21
Compreh.–Pict. Comp.	0.66	0.05
Compreh.–Pict. Arr.	2.27	1.18
Compreh.–Object A.	0.77	0.17
Compreh.–Coding	2.73	0.69
Coding–Digit Span	4.95	2.75

Differences which are statistically significant at the 5 per cent level, or higher, occurred for the following intertest comparisons within the experimental group: Comprehension–Vocabulary, Comprehension–Coding, Digit Span–Comprehension, and Comprehension–Picture Arrangement. For the control group only two such intertest comparisons are significant at the 5 per cent level, or higher: Comprehension–Vocabulary and Digit Span–Comprehension.

For the experimental group, but not for the control subjects, the Coding test proved to be significantly inferior to the Comprehension test. It is possible that the experimental cases experienced special difficulty in attempting to use paper and pencil, a possibility suggested by Wechsler [10, p. 49]. Or, the syphilitic cases may lack, or have lost, the ability to respond effectively to the demands of a new learning situation. It is assumed that the learning ability of such subjects would show progressive impairment unless adequate treatment were available.

For the syphilitic group the significantly poor record on the Picture Arrangement test, as compared with the Comprehension test, reveals a relative lack, or loss, of the ability to comprehend part-whole relationships, or to appraise total situations.

It is of interest to observe that negligible differences result from comparisons of the syphilitic's records on the Comprehension test with each of the following tests: Object Assembly, Picture Completion, Similarities, Arithmetic, and Information.

It is apparent from the results of this study that there is need for further use of the WISC with large numbers of Negro children before interpretations of scores of special groups can be accurately given.

SUMMARY

The Wechsler Intelligence Scale for Children was given to 40 Negro congenital syphi-

litic children and 40 control cases. From an analysis of the data the following conclusions seem justified:

The syphilitic group had a mean IQ of 59 and the control group a mean IQ of 70. According to Wechsler, these subjects would be classified as "mentally defective" and "borderline," respectively. According to socioeconomic or observational criteria the control group would not be regarded as "borderline." It appears that Wechsler's norms obtained on white children, are inapplicable to a sample of rural southern Negro children.

The subjects in the present study had better Verbal than Performance IQ's, which is contrary to the usual findings for rural cases.

The control subjects had better scores on all eleven subtests, and all but three comparisons of subtest scores resulted in differences which are significant. Deficient learning ability, inferior generalizing power, and difficulty in dealing with part-whole relationships characterize the syphilitic group.

Using the Comprehension test to represent a relatively stable function, the following intertest comparisons within the syphilitic group yielded significant differences: Comprehension–Vocabulary, Comprehension–Coding, Digit Span–Comprehension, and Comprehension–Picture Arrangement. For the control group, the Coding and Picture Arrangement tests did not differ significantly from the Comprehension test; therefore, the inferiority of the syphilitics in learning new material and in comprehending part-whole relationships is noteworthy.

Congenital syphilis in the children of the present study is associated with significantly inferior intelligence and with selective impairment of mental functions.

Received September 6, 1950.

REFERENCES

1. BABCOCK, HARRIET. An experiment in the measurement of mental deterioration. *Arch. Psychol., N. Y.*, 1930, No. 117.
2. BISCH, L. E. *Clinical psychology.* Baltimore: Williams & Wilkins, 1925.
3. FISHER, V. E. *An introduction to abnormal psychology.* New York: Macmillan, 1937.
4. JENKINS, L. R., BROWN, A. W., AND CISLER, L. E. The influence of syphilis on the intelligence of children. *Amer. J. Dis. Child.*, 1940, **60**, 341-352.
5. LANDIS, C., AND BOLLES, M. MARJORIE. *Abnormal psychology.* (Rev. Ed.) New York: Macmillan, 1950.
6. SEASHORE, H. G. Differences between verbal and performance IQ's on the Wechsler Intelligence Scale for Children. *Amer. Psychologist*, 1950, **5**, 319. (Abstract.)
7. SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, **9**, 371-377.
8. STOKES, J. *The third great plague.* Philadelphia: Saunders, 1937.
9. STRECKER, E. A., EBAUGH, F. G., AND EWALT, J. R. *Practical clinical psychiatry.* Philadelphia: Blakiston, 1937.
10. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
11. WECHSLER, D. *Wechsler intelligence scale for children.* New York: Psychological Corp., 1949.

THE RELATION OF WRITING SPEED TO AGE AND TO THE SENILE PSYCHOSES[1]

JAMES E. BIRREN AND JACK BOTWINICK

NATIONAL HEART INSTITUTE, NATIONAL INSTITUTES OF HEALTH, BETHESDA, MARYLAND AND THE GERONTOLOGY SECTION, BALTIMORE CITY HOSPITALS, BALTIMORE, MARYLAND

THE purposes of this study were: (*a*) to determine the relation between speed of writing and chronological age in individuals of a wide age range, and (*b*) to determine the relation of writing speed and the senile psychoses. One of the most distinctive features of the elderly is a slowness of speed of response [6], and it was thought important to investigate loss of speed in a highly overlearned function, e.g., writing. There has been no report in the literature of a specific attempt to study age differences in speed of handwriting in normal persons, although tests of writing and figure copying have been used in the measurement of intellectual deterioration [1]. It is known that senile brain changes are reflected in qualitative disturbances of handwriting [5, pp. 60-64], and patients with dementia paralytica have been shown to have a marked loss in speed of writing [8]. It seems reasonable to expect that individuals with senile psychosis or psychosis with cerebral arteriosclerosis might also show a loss of speed of writing.

Tests of writing speed appeared to be potentially useful in the psychological evaluation of the elderly and offered several desirable features: (*a*) the material is simple and familiar, (*b*) the responses required (writing) are highly overlearned, (*c*) no special apparatus is required, and (*d*) the test series may be easily lengthened to secure any degree of reliability desired.

PROCEDURE

Materials. Pages of digits were prepared by selecting digits from tables of random numbers. These digits were arranged in eight columns and 12 rows per page, a total of 96 digits per page. The columns were one inch apart and the rows three-quarters of an inch apart. The test materials were duplicated from stencils typed with primer-size type, eight characters per inch. The wide spacing and large characters were used to reduce difficulties of legibility for the elderly.

Pages of random words were typed in a similar manner. The words were selected from the text of a daily newspaper, eliminating unusual words, proper nouns, and articles. The word length ranged from four to ten letters, with a mean of six. The words were arranged in three columns of 23 words each, the first letters of each word between columns were two and a half inches apart. The words in a column were three-eighths of an inch apart. The words were randomized, and there was no apparent context, e.g., the first ten words were "area, today, work, called, swirl, southern, where, ground, hungry, horse."

Test Administration. The tests were given both in group fashion and individually; the latter method was used for persons over sixty years. In group testing the time limits per page were 40 seconds for digits and two minutes for words. In individual administration of the digits test, the time limit was either two minutes or time to complete the page, whichever came first. The time limit on both individually and group administered word tests

[1]The cooperation of the staff of the Baltimore City Schools, the cooperation of the staffs of the Spring Grove and the Springfield Maryland State Hospitals and the Sheppard and Enoch Pratt Hospital, the assistance of Miss Charlotte Fox in gathering some of the data, and the computational assistance of Mrs. Betty Benser are gratefully acknowledged.

was two minutes. In all instances the subjects were told that the object of the test was to see how fast they could write. All writing was done with pencil. Subjects were told to write in their usual handwriting: "Write these words as fast as you can but be sure I can read your handwriting. Write as you usually do; don't try to be fancy. When I say 'Go,' write the words as fast as you can."

Because of the different time limits, speed of writing digits was expressed as digits per second and derived simply by dividing the number of digits by the total writing time in seconds. Speed of writing words was expressed as the total number written in two minutes.

Subjects. A total of 554 subjects aged 16 to 89 years was used in this study. All subjects were white, born in the United States or England, and had a minimum of four years of schooling. Four general groups of subjects were used: high school seniors; adults taking evening school work in crafts and arts; the professional, technical, and custodial staffs of the Baltimore City Hospitals; and the residents of the Baltimore City Hospital home for the indigent aged. The high school students were selected by classes to give an average group. The high school students' mean IQ, obtained by the Otis Self Administering Test of Mental Ability (Form A), was 104, with a range of 83 to 135 and σ of 9.3.

In addition to the above subjects, a total of 35 patients diagnosed as senile psychosis or psychosis with cerebral arteriosclerosis was studied. The patients were selected from the populations of three mental hospitals after detailed screening involving a review of the case histories, discussions with the physicians in charge, and an interview with each patient [2]. All patients were ambulatory and were able to sit at a desk and write. Only patients in whom cerebral arteriosclerosis or senile brain changes were regarded as playing a dominant role in the syndrome were included. Patients having previous admissions to a mental hospital in early or middle life were excluded. The mean age at time of first admission was 61 years, with a range of 51 to 68 years. The patients were all in the age range 60 to 70 years inclusive and met the same criteria as the general population, i.e., they were white, their native language was English, and they had a minimum of four years of schooling. The senile patients and the general population of the same age were approximately equated for years of education: mean years of education, Senile—8.1, Control—7.8; median, Senile—7.8, Control—8.1; range, Senile—4 to 17, Control—4 to 16; σ, Senile—3.9, Control—3.0.

There were relatively more individuals with college training in the senile groups, five out of 35, than in the normal or control group, six out of 68. Thus, if any slight bias exists in the sample, it is in the direction of having a few more individuals with superior education in the senile group.

Reliability. Several estimates of the reliability of the digits test were obtained. In each instance the split-half reliability was determined by correlating (r) the results of alternate pages of random digits using a 40-second time limit for each sheet. The reliability of a 40-second test was about .85 (Table 1). In practice, the results of two tests of 40 seconds each were combined for the individual. The reliability of this longer measurement is about 0.92 when corrected for test length [4, p. 419]. Thus, one may obtain satisfactory reliability even with short time limits on the test, for groups homogeneous in education, age, sex, and nativity. The two-minute test used individually with the elderly adults would be higher than that obtained with shorter time limits in group testing, i.e., about .94 when corrected for test length (4, p. 419).

No direct check of the reliability of the two-minute word writing test was made with alternate forms. This time limit is satisfactory, however, as the correlation between the

TABLE 1

CORRELATIONS, UNCORRECTED FOR LENGTH, BETWEEN ALTERNATE FORMS OF A 40-SECOND TEST OF SPEED OF WRITING DIGITS

Group	N	Coefficient of Reliability (r)
Senior high school girls	73	.84
Senior high school boys	64	.85
Evening school adults	70	.86

words and digits tests in a group of 63 adults aged 60-69 years was .86, and the correlation for all subjects aged 16 to 89 was .93 ($N = 536$). The reliability of the word test would be expected to be at least the lower of these two values.

RESULTS

Speed of writing both digits and words was significantly slower in the older age groups (Table 2), and there was a further reduction in writing speed in the patients with senile psychoses (Table 3).

The r's between age and speed of writing digits and words in the total groups were —.81 and —.83, respectively; the corresponding η's were —.83 and —.87. If one considers only the cases aged 20 and older, only a slight curvilinearity is present in the relation of digits to age, and no significant curvilinearity in the relation of words to age. The r's for subjects aged 20 and older were —.78 for digits and —.85 for words; the corresponding η's were —.80 and —.85.

There was little overlapping in the distributions of scores of the high school students and the subjects in their seventies and eighties (Figure 1). The distributions of the scores of the high school students and the aged appear more nearly symmetrical than the middle-aged groups.

The two tests, digits and words, correlated .93 for all age groups ($N = 536$). The partial correlation [4, p. 399] between digits and words for all subjects, removing the covariance of age, was .76. The partial correlation, however, does not take into account the slight curvilinearity in the data. An estimate of the same correlation is obtained by averaging the correlations between digits and words within each decade, i.e., after a z transformation. The correlation thus obtained was .78, which approximates the partial r of .76 just cited.

Since the level of educational attainment has been increasing [10], there is an inherent correlation of age and education in all random sampling. In the present study the effect of education was removed by a partial correlation. The correlation of age with education was —.65, $N = 369$, for all subjects 20 years and older in the present sample. When this relation was extracted from the correlations between age and both digits and words, the partial correlations were —.65 and —.68, respectively. It seems reasonable to regard these correlations as more representative of the relation of speed of writing and age, than those obtained when level of education is permitted to vary.

When the years of education was correlated with speed of writing, removing the covariance of age by a partial correlation, it is shown that educational attainment affects speed of writing. Thus the partial correlations were .48 and .61 between education and digits, and education and words, respectively. It would appear from these results that speed of writing digits is a better measure of speed of response in the elderly, as it is somewhat less influenced by education than is speed of writing words. If one removes the effects of both age and education from the correlation of

TABLE 2

SPEED OF WRITING DIGITS AND WORDS IN SUBJECTS 16 TO 89 YEARS OF AGE

Age:	16–20*	20–29	30–39	40–49	50–59	60–69	70–79	80–89
				DIGITS PER SECOND				
M	1.67	1.62	1.63	1.29	1.03	0.82	0.57	0.49
σ	0.24	0.26	0.39	0.40	0.41	0.32	0.20	0.20
N	189	57	53	49	41	66	55	37
				WORDS PER TWO MINUTES				
M	45.2	48.0	47.1	38.0	29.4	22.0	16.5	15.2
σ	5.2	6.6	9.7	12.0	12.0	9.7	7.3	7.0
N	190	60	53	52	42	68	55	34

*High school students.

writing speed of digits and of words, this second-order partial correlation is .70. It is to be remembered that this partial correlation assumes linearity in all the zero-order and first-order partial correlations. Although the data indicated some curvilinearity, the obtained correlation is an acceptable approximation, since the curvilinearity is slight.

When the patients with senile psychoses were compared with the general population of the same age, the senile patients were seen to be significantly slower for both forms of writing (Table 3).

TABLE 3

DIFFERENCES IN WRITING SPEED BETWEEN SENILE PATIENTS AND CONTROL SUBJECTS IN THE AGE RANGE 60 TO 70 YEARS INCLUSIVE

	Mean	σ	N	Diff	σ_{diff}	t	P
Words/2 min.							
Senile	16.1	9.2	35	5.9	1.97	2.99	<.01
Control	22.0	9.7	68				
Digits/sec.							
Senile	0.52	0.22	35	0.30	0.054	5.56	<.01
Control	0.82	0.32	66				

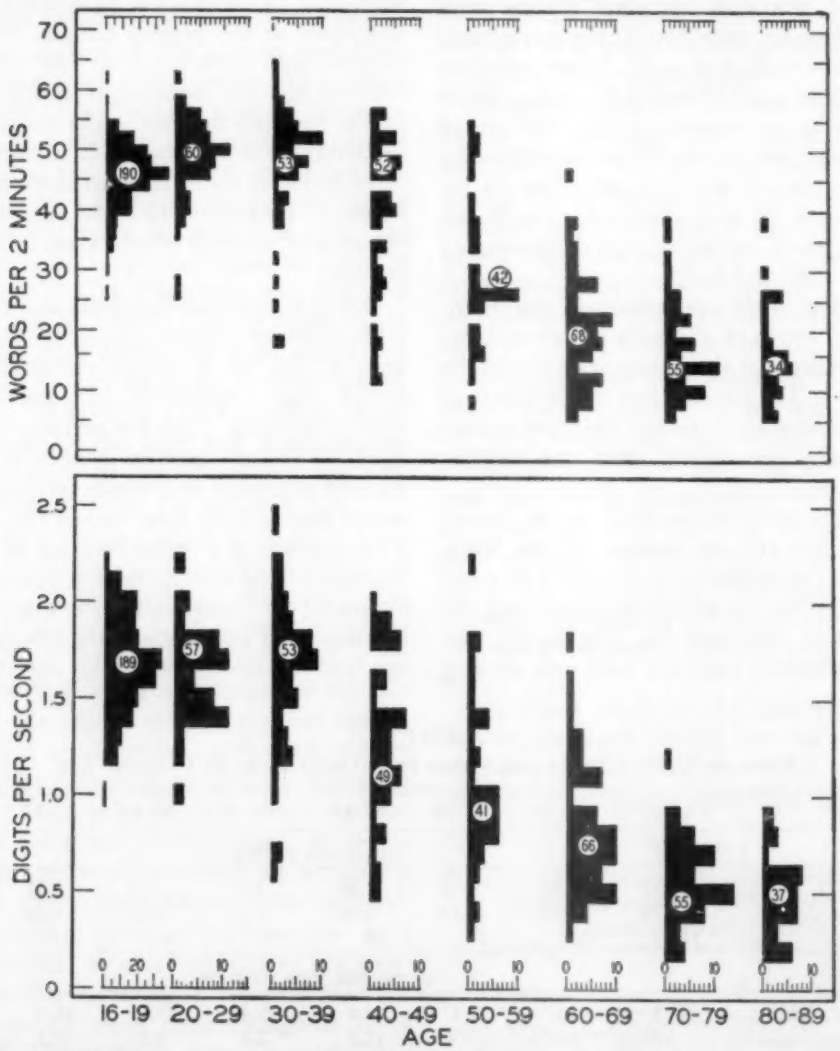

FIG. 1. Age differences in writing speed. Frequency polygons are shown for each age group. The total number of subjects in each age group is indicated on the individual distributions; the number of individuals in a single interval may be estimated from the frequency scales on the abscissa.

The t test of significance does not reflect the true nature of the differences between the senile and general groups, since there is a tendency toward skewing in the distributions, and the σ's are unusually large. The nature of the distributions may be appreciated in Figure 2, wherein the individual values have been plotted. The individual values of a group of 20- to 29-year-old normal subjects are also plotted to demonstrate the age difference as well as the difference in mental status. It can be seen that the difference between the 20- and 60-year-old normal individuals is greater than the difference between the 60-year-old normal and senile individuals. The difference between the young and aged subjects is, of course, enhanced somewhat by the current trend toward increasing education [10]. The larger decrement with age as compared with the decrement in the senile psychoses has been seen on other tests [2] and clearly suggests the necessity of having control groups carefully matched for age with experimental or patient populations.

DISCUSSION

The age differences in speed of writing are large when compared with changes in other

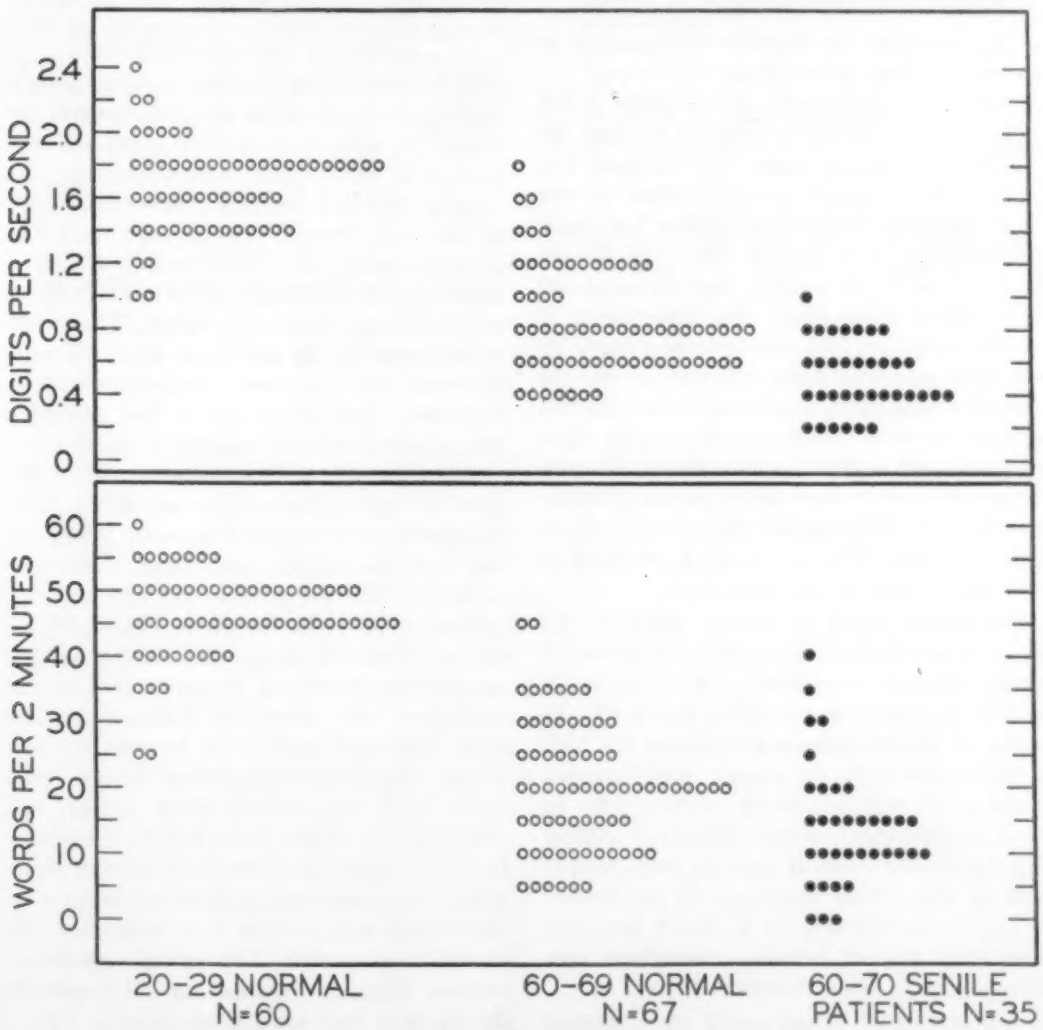

FIG. 2. Speed of writing in patients with senile psychoses compared with control subjects in two age groups. Individual scores are plotted in frequency distributions; O = control subject, • = senile patient.

psychological and psychophysiological functions [6]. In addition, the present results show a further reduction in writing speed in patients with senile psychoses. Loss of writing speed may result from several factors, e.g., slow visual perception, disuse changes in the muscle groups used in writing, arthritic involvement of the fingers, and changes in the brain as reflected in aphasic and apraxic signs. The nature of the disturbance may to some extent be determined by noting errors in reading as well as in writing. The difference in speed of reading and writing may indicate something of the general nature of the slow performance. In the very deteriorated, errors occur in naming the digits or words as well as in reproduction by writing.

There are individuals in the senile group whose writing of digits appeared to have the quality of drawing rather than writing, i.e., they required many revisualizations of the digit and they broke the response into component lines. It is only by observing the patient in the act of writing that the examiner may obtain clues about the disturbance in writing [3; 7, p. 91]. Because the slowest individuals appeared more prone to errors, the tentative hypothesis is advanced that the loss of speed in such simple tasks as writing digits or words and in reading them aloud, is a precursor to, and in later stages occurs concomitantly with, characteristic disturbances of an aphasic nature. Further work is required to test the validity of this hypothesis.

An earlier study of writing speed in 172 primary school children, aged eight to twelve years, reported a writing rate of about 24 words per minute in the oldest group [9]. In terms of letters written per minute the high school students of the present study showed about a 13 per cent faster writing rate. In view of differences in the methods of obtaining the measurement, it does not seem reasonable to place much confidence in the magnitude of this increase. It is likely, however, that writing speed increases throughout adolescence, as do other functions [6].

Limitations of visual acuity often present in the elderly were minimized in the present study by the use of reading glasses and by the use of large type size. To remove the artifact of illiteracy, all subjects with less than four years of schooling were excluded. Level of education (and perhaps occupation) should be considered in evaluating an individual's performance. When these factors are taken into consideration, measurement of writing speed may prove to be a valuable adjunct to the psychological examination of the elderly person in whom deterioration is suspected. The measurement of writing speed is reliable and is simple for all but the most deteriorated patients. The test requires little time to administer. These features make it a desirable measure for use with the elderly.

SUMMARY AND CONCLUSIONS

1. Measurements of speed of writing digits and words were made on 554 subjects between the ages of 16 and 89 years. All subjects were white, their native language was English, and they had had four or more years of schooling. Persons over 60 years were tested individually and with reading glasses if necessary. In addition, a total of 35 patients with senile psychoses was studied. These individuals were in the age range 60 to 70 years inclusive and had been institutionalized in later life. Their mean age at first admission was 61 years, with a range of 51 to 68 years.

2. Large age differences were found for speed of writing both digits and words. The correlation between speed of writing and age was high and slightly curvilinear in the case of digits. The η's between age and speed of writing were —.83 for digits and —.87 for words. When the educational level of all age samples was controlled by means of a partial correlation, the correlation between age and digits was —.65 and —.68 between age and words. The partial correlation between education level and writing speed, holding age constant, was .48 for digits and .61 for words. It would appear that speed of writing digits is the better measure of speed of response in the elderly inasmuch as it is somewhat less related to education. The partial correlation between digits and words was .70, removing the effects of both age and education.

3. Significant differences were noted between the patients with senile psychoses and

control subjects of the same age and education. The speed of writing of both digits and words was slower in the senile patients; the mean differences amount to about —.06 σ for words and —.09 σ for digits, using the σ of control group as the base. Both differences were statistically significant. The tentative hypothesis is advanced that the loss of speed in senile patients is a precursor to aphasic disturbances. The difference in writing speed between the senile patients and the normal elderly was less than the difference between the normal elderly and young subjects.

4. The loss of writing speed in later life and the further reduction of writing speed in the senile psychoses, and the reliability and ease of making the measurements, indicate that tests of writing speed may be a simple and valuable adjunct to the psychological examination of the older person.

Received August 14, 1950.

References

1. Babcock, Harriet. An experiment in the measurement of mental deterioration. *Arch. Psychol., N.Y.*, 1930, No. 117.
2. Botwinick, J., and Birren, J. E. The measurement of intellectual decline in the senile psychoses. *J. consult. Psychol.*, 1951, **15**, 145-153.
3. Eaton, H. D. Handwriting, a neurological study. *Calif. West. Med.*, 1938, **48**, 430-435.
4. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
5. Koster, R. *Die Schrift bei Geisteskrankheiten.* Leipzig: Barth, 1903.
6. Miles, W. R. Psychological aspects of ageing. In E. V. Cowdry (Ed.), *Problems of ageing* (2nd Ed.). Baltimore: Williams & Wilkins, 1942. Pp. 756-784.
7. Nielsen, J. M. *Agnosia, apraxia, aphasia, their value in cerebral localization.* New York: Hoeber, 1947.
8. Quinan, C. The time and lineage factors of handwriting: An experimental study. *Arch. Neurol. Psychiat., Chicago*, 1931, **26**, 333-345.
9. Wills, A. R. An investigation of the relationship between rate and quality of handwriting in primary school. *Brit. J. educ. Psychol.*, 1938, **8**, 229-236.
10. Educational characteristics of the population of the United States, by age: 1940. Washington, D. C.: Bureau of the Census, Series P-19, No. 4, 1943.

INTERNAL CONSISTENCY OF THE GROUP CONFORMITY RATING OF THE ROSENZWEIG PICTURE-FRUSTRATION STUDY[1]

MAHLON V. TAYLOR, JR., AND O. M. TAYLOR
ALLEGHENY VOCATIONAL COUNSELING CENTER
PITTSBURGH, PENNSYLVANIA

THE Rosenzweig Picture-Frustration Study (referred to herein as the P-F) yields, among other measures, a Group Conformity Rating (GCR) [10, p. 201]. Rosenzweig states that the GCR "forms one basis for judging social adjustment" [11, p. 168], is indicative of "capacity for conformity with conventional behavior under stress" [11, p. 172], and may reveal whether a subject is "capable of conventional responses in social situations and . . . probably . . . , at least superficially, better adjusted" [11, p. 180].

Capacity or capability, of course, can only be inferred, since neither is necessarily ever revealed in overt behavior. Rosenzweig's concept of "group conformity" may be said to be at the level of a hypothetical construct, and subject to experimental verification against an external criterion, but not subject to disproof by the same method. At least one corollary can, however, be deduced from the concept, and furnish a hypothesis which can be tested against a criterion internal to the GCR.

For those unfamiliar with the mechanics of the P-F and GCR, a brief description, paraphrased from the manual [10, pp. 165-166, 201], may serve to clarify some of the arguments to follow.

Description of the GCR. The items are cartoon-like pictures, each depicting two persons involved in a mildly frustrating situation of common occurrence. One person is shown saying certain words which either help to describe the frustration of the other individual, or which are themselves actually frustrating to him. The subject is instructed to write the first appropriate reply which enters his mind.

[1]This study was conducted at the Allegheny Vocational Counseling Center with the assistance of a grant from the Buhl Foundation.

The GCR score is the percentage of responses which agree with the criterion scores on 12 items previously found to elicit a particular variety of response from normal subjects significantly often.

From this description it is seen that the items selected for the GCR, from the 24 comprising the P-F, are those eliciting significantly popular responses. It is these responses which are given positive scoring weights of unity (or a fraction in the case of combination responses), while less popular responses are weighted zero. Thus the larger the proportion of popular responses given by a subject, the higher his score for "group conformity."

A measure operationally defined in this way assumes the existence of a trait of group conformity. Cattell [2, p. 61] defines a trait as "a collection of reactions or responses bound by some kind of unity which permits them to be gathered under one term and treated in the same fashion for most purposes." The unifying element in this case seems to be the tendency of subjects to make popular or conventional responses. All that is needed to demonstrate the reality of the trait is to show that subjects who tend to make conventional responses in some frustrating situations also tend in similar degree to make conventional responses in other similarly frustrating situations.

Lacking observations of reactions in real-life situations, the demonstration can be attempted with the GCR items, and may be considered successful, although not to be generalized to real-life situations, if the item scores are found to be positively intercorrelated. This amounts to testing the hypothesis of internal consistency of the GCR. The higher

the positive intercorrelations among the items, the greater the internal consistency, and the greater the average item saturation with a general factor of "group conformity." Thus determining the internal consistency of the GCR may at one time furnish evidence on the reality of a trait of group conformity, and assess the reliability of the GCR as a measure of the trait. A survey of the literature does not bring to light any investigation of this kind.

METHOD

The preferable method for estimating internal consistency of the GCR appears to be analysis of variance. Hoyt [6, pp. 155, 158] points out that it "gives a better estimate than any method based upon an arbitrary division of the test into halves," and also "gives precisely the same result as formula (20)" of Kuder-Richardson [9, p. 158].

The total variance of responses is apportioned among items, individuals, and remainder (or error). The coefficient of reliability, in the sense of internal consistency rather than test-retest [7, p. 157], is expressed as:

$$r_{tt} = \frac{\text{(variance among individuals)} - \text{(remainder variance)}}{\text{(variance among individuals)}}.$$

This is the ratio of estimated true variance among individuals (i.e., obtained variance among individuals minus remainder or error variance) to the obtained variance among individuals. The resulting value is "the percentage of obtained variance that is not associated with unreliability of the test . . . " [6, p. 160]. In correlational terminology, the numerator term of r_{tt} represents the portion of response scores which is correlated among individuals, since the subtracted error variance is the uncorrelated portion attributable to chance fluctuation, again emphasizing the dependence of internal consistency upon positive item intercorrelations [9, p. 159].

THE STUDY

The data were provided by P-F records for 130 adult males tested for purposes of personnel evaluation or vocational guidance. Responses had been carefully scored by Rosenzweig's criteria [10, pp. 201, 203], and doubtful scorings independently checked. GCR response scores were tabulated by individual and item. Eighteen of 1,560 responses were found to be omitted or unscorable. Adjusted values were computed for the missing responses [12, p. 268], then taken to the nearest obtained response value (i.e., 0.0, 0.5, or 1.0), and these values inserted in the appropriate blank cells, with degrees of freedom for total and remainder reduced by 18.

TABLE 1
ANALYSIS OF VARIANCE OF THE 12 P-F GCR ITEMS ADMINISTERED TO 130 ADULT MALES

Source of Variation	*d.f.*	Sum of Squares	Variance
Total	1,541	206.7512	0.1342
Among items	11	44.4363	4.0397
Among individuals	129	18.2546	0.1415
Remainder	1,401	206.7512	0.1476

The results of the analysis of variance are shown in Table 1. The reliability coefficient is

$$r_{tt} = (.1415 - .1476)/.1415 = -.0431.$$

The variance among individuals is slightly less than the remainder or error variance ($F = .1415 / .1476 = .9587$), while variance among items is highly significant ($F = 4.0397 / .1476 = 27.3692$).

While the device of inserting adjusted values for missing scores seemed preferable, the results with two less defensible treatments were also calculated. Leaving 18 cells blank and reducing degrees of freedom by 18 for total and remainder gave

$$r_{tt} = \frac{.1500 - .1509}{.1500} = -.0060.$$

Treating the blank cells as zero response scores, i.e., involving no loss in degrees of freedom, gave

$$r_{tt} = \frac{.1377 - .1474}{.1377} = -.0704.$$

Thus any of the three treatments of the data would lead to the same conclusions.

DISCUSSION AND CONCLUSIONS

It is plain from the obtained reliability coefficient ($r_{tt} = -.04$) that the present data do not support the hypothesis of internal consistency of the GCR. There is no demonstration of the reality of a trait of "group conformity," nor is the GCR shown to be a reliable measure of any trait. Stated in another way, there is no evidence of a general factor arising from positive intercorrelations among the items.

It may be noted that Bernard [1], after

basing GCR scores on his sample data, but using Rosenzweig's criteria for selecting items and scoring responses, obtained test-retest reliability of only .45 with a three- to nine-month interval. Thus it appears from Bernard's [1] and the present study that P-F items, selected and scored by Rosenzweig's scheme for the GCR, should not be expected to measure "group conformity" with satisfactory reliability, either among different individuals at the same time, or for the same individual at different times.

Technical Discussion. Tests so seldom yield negative estimates of internal consistency that the obtained r_{tt} of —.04 may be viewed with suspicion by some. However, it is pointed out by Jackson and Ferguson [8, p. 75] that a test's average inter-item correlation may take a negative value. Also Hayes and Clark [5, p. 7*a*, Table 4] report a negative estimate, and Clark [3] states that it "reflects, perhaps, an idiosyncracy of Kuder-Richardson formula 20." With the present method it is clear that a negative estimate is obtained whenever variance among individuals happens to be smaller than the remainder, or error, variance. The probability of such a condition arising in a sample is very slight if the population value of the reliability coefficient is considerably different from zero.

Although not directly pertinent to the present study, it is perhaps worthy of comment that the very significant variance among items is attributable to widely varying popularity among items of the responses scored as conventional. Gulliksen shows that "as the difficulty difference between two items increases, the maximum possible correlation between them drops sharply" [4, pp. 81–82], and "to maximize the reliability and variance of the test, items should have high intercorrelations, all items should be of the same difficulty level, and this level should be as near to 50% as possible" [4, p. 90]. These conclusions are, of course, applicable to measures like the GCR if "popularity" is substituted for "difficulty." Thus, even if the GCR were measuring a trait or general factor, its reliability would be attenuated by the considerable dispersion among items of popularity levels of the "conventional" or scored responses.

SUMMARY

Rosenzweig's interpretation of the Picture Frustration Study Group Conformity Rating as indicating "capacity for conformity with conventional behavior under stress" was said to imply the reality of a trait of "group conformity." Evidence for existence of the trait and the reliability (i. e., internal consistency) of the GCR was sought by subjecting the GCR response scores of 130 adult males to analysis of variance.

Variance among items was found to be very significant ($F = 27.37$) but variance among individuals negligible—in fact slightly less than the remainder, or error, variance ($F = .96$). Thus, the reliability coefficient was —.04. Technical implications of these findings were discussed.

It was concluded that the present data neither provide evidence that a trait of "group conformity" exists, nor that the GCR is a reliable measure of any trait.

Received October 12, 1950.

REFERENCES

1. BERNARD, J. The Rosenzweig Picture-Frustration Study: I. Norms, reliability, and statistical evaluation. *J. Psychol.*, 1949, **28**, 325-332.
2. CATTELL, R. B. *Description and measurement of personality.* Yonkers, N. Y.: World Book Co., 1946.
3. CLARK, K. E. Private communication. August 21, 1950.
4. GULLIKSEN, H. The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 1945, **10**, 79-91.
5. HAYES, PATRICIA M., AND CLARK, K. E. Maximizing differences in vocational interest patterns of occupational groups by varying methods of scoring. Dept. of Psychology, Univ. Minn., Technical Report No. 3, October, 1949.
6. HOYT, C. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153-160.
7. JACKSON, R. W. B. Note on the relationship between internal consistency and test-retest estimates of the reliability of a test. *Psychometrika*, 1942, **7**, 157-164.
8. JACKSON, R. W. B., AND FERGUSON, G. A. Studies on the reliability of tests. Dept. of Educational Research, Univ. Toronto, Bulletin No. 12, 1941.
9. KUDER, G. F., AND RICHARDSON, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151-160.
10. ROSENZWEIG, S., FLEMING, EDITH E., AND CLARKE, HELEN JANE. Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, **24**, 165-208.
11. ROSENZWEIG, S., AND KOGAN, KATE. *Psychodiagnosis.* New York: Grune & Stratton, 1949.
12. SNEDECOR, G. W. *Statistical methods.* (4th Ed.) Ames, Iowa: Iowa State Coll. Press, 1946.

MMPI PROFILES IN MULTIPLE SCLEROSIS[1]

AARON H. CANTER
VETERANS ADMINISTRATION MENTAL HYGIENE CLINIC,
PHOENIX, ARIZONA

A CAREFUL examination of the American literature on multiple sclerosis, as Sugar and Nadell [5] have recently pointed out, reveals a paucity of material dealing with the mental or emotional symptoms of that disease. The meeting of the Association for Research in Mental and Nervous Disease in December, 1948, which was devoted to a discussion of the various aspects of the disease, served to emphasize not only our relative lack of information about multiple sclerosis but, particularly, the need for research studies which might add to our understanding and ultimately to its control.

The present report of research is part of a more comprehensive and continuing investigation begun in early 1948 at the Veterans Administration Mental Hygiene Clinic, Brooklyn Regional Office, and which was described in a previous publication by Canter [1]. The purpose of the present report is to study the psychological aspects of multiple sclerosis as revealed by the Minnesota Multiphasic Personality Inventory (MMPI) in order to determine whether the data reveal any reliable personality patterns associated with the disease.

The MMPI [3] is a widely-used, structured test on which personality characteristics are described by a profile consisting of scores on nine component personality scales: hypochondriasis (*Hs*), symptomatic depression (*D*), hysteria (*Hy*), psychopathic deviate (*Pd*), masculinity-femininity (*Mf*), paranoia (*Pa*), psychasthenia (*Pt*), schizophrenia (*Sc*), and hypomania (*Ma*). In addition, four "validity" scales, "*?*," "*L*," "*F*," and "*K*" are included to indicate whether a given test record may be depended upon as a description of the patient's personality, or whether his indecision, attitude towards personality tests, atypicality, or other similar tendencies make the record invalid. Meehl [4] has suggested a method of configural scoring which should further extend the value of the MMPI in clinical work. He furnishes a bibliography of the more recent research studies utilizing the scales.

Many previous studies of the MMPI have been concerned with the demonstration of statistically significant differences between various diagnostic groups. In these studies, patterns characteristic of each group have been obtained by constructing a profile from the average standard scores found on each of the nine personality scales. Admittedly, these "average profiles" have limited clinical value in the treatment and understanding of the individual patient. Descriptive studies of multiple sclerosis, however, must first be made prior to attempting cross-validation studies, or studies which can be of direct value in the treatment of individual patients. As was previously indicated, this present report is concerned primarily with identification of the personality patterns associated with multiple sclerosis.

SUBJECTS

The subjects of this study are 33 of the patients comprising the experimental group of multiple sclerotics described in detail in a previous study [1]. In general, the group constitutes a fairly large and representative sample of the World War II veterans in the early stages of the disease process. The mean age of the group was approximately 32, and the duration of the acute symptoms ranged from one to eight years, with an approximate average of four years. In all the cases sufficient organic damage exists to make the medical diagnosis of multiple sclerosis certain.

[1]Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

PROCEDURE

The group form of the MMPI was administered as part of a battery of psychological tests in the preliminary phases of the study of multiple sclerotics. The patients were informed at the outset of the initial testing that the purpose of the psychological examinations was to secure data on multiple sclerosis which might contribute to a better understanding of the disease. As previously described, the experimental group revealed a maximum amount of motivation and was uniformly cooperative in complying with the testing procedure. Of the 33 patients who took the MMPI initially, 21 patients received a retest within one week of the date exactly six months later. Twelve members of the experimental group did not receive the retest at the exact six-month date because of administrative difficulties. There is no reason to believe that the cases for which retests are available differ significantly from the entire group of cases.

To increase the value of the objective data, a clinical team of a neuropsychiatrist, a clinical psychologist, and a psychiatric social worker[2] were asked to write independently a "brief impression" gained from the inspection of the average profile of the 33 multiple sclerotics. These impressions were not made with the purpose of attempting a complete personality description or evaluation, but simply to indicate what appeared (to the judges) to be the significant personality trends revealed by the profiles.

RESULTS

The data necessary for answering the problems of this investigation are set forth in Table 1 and Figure 1. Records were, of course, examined and analyzed individually, but the material has been combined into one profile for simplicity of presentation. This procedure involves obscuring qualitative differences, but because of relatively narrow ranges of scores on the component personality scales, the technique is not seriously in error.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF MMPI CATEGORIES

MMPI Category	Initial Test ($N = 33$) M*	SD	Retest ($N = 21$) M*	SD
?	50.4	1.9	50.4	1.9
L	54.5	3.9	52.6	3.7
K	57.8	8.4	54.2	8.4
F	55.7	5.7	53.3	5.0
Hs	81.2	12.9	79.0	12.7
D	79.4	12.8	78.8	12.7
Hy	75.3	10.0	76.6	9.3
Pd	59.2	11.5	59.0	10.2
Mf	55.4	8.7	53.5	9.1
Pa	53.2	9.2	51.6	9.2
Pt	63.0	16.0	61.4	14.9
Sc	63.6	14.4	62.1	14.1
Ma	55.1	9.2	52.1	9.3

*Values are reported in T scores.

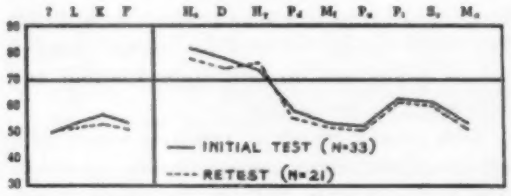

FIG. 1. MMPI profiles of the multiple sclerotic groups.

A comparison of the two profiles reveals that their general shape is quite similar. There is a small, but general, regression of the mean scores on the component scales in the direction of the general population norms, which might be attributed to the slight clinical improvement of the group as a whole in their acceptance and adjustment to the disease process. However, this trend revealed by the retest is more apparent than real. A statistical analysis (the t test of the difference in means) fails to reveal any significant changes of the means in the component scales upon re-examination.

Excerpts of the impressions of the significant personality trends noted by the "clinical team" working independently were as follows:

The MMPI curve is an over-all mildly elevated curve indicating a pervasive maladjustment. The *Hs* and *D*, as the highest scores, suggest preoccupation with physical functioning is tinged with

[2]The author expresses indebtedness to Dr. Kenneth G. Rew, Neuropsychiatrist, Dr. William Hunter, Psychologist, and Mrs. Greta Mann, Psychiatric Social Worker, all of the Regional Office, Phoenix, Arizona, for their assistance in this phase of the study.

hopelessness. With the *Hs* higher than the *Hy* score, the indication is that physical preoccupation is a relatively malignant one and is not balanced by tension-relieving dependence or conversion. The slight elevation of the *Pt* and *Sc* scores indicates ego-disorganizing processes and may be a warning of future development.

This looks like the curve of a person reacting to situational stress such as injury or chronic disease. It would seem that the *Pt* and *Sc* elevations must represent an attempt to deal with anxiety centering around conflicts having to do with body image and capacities (*Hs*). The *D*, which is not significantly high, suggests a feeling of apathy and hopelessness rather than a true clinical depression. This would suggest that deep-seated dependency needs do not seem a factor, and a retreat into invalidism is not an acceptable way out. The *Hy* is not sufficiently elevated to suggest a conversion solution. There is apparently much free-floating anxiety about self and health that the ego has to handle. The significant aspect of this curve for therapy is the high *Pt* and *Sc*, which reveal a way of handling anxiety that may threaten the ego sufficiently to precipitate a psychotic break.

There is apprehension over health, depression, some irritability and aggressiveness leading to some difficulty in social adjustment. There is a tendency to act immaturely and unrealistically with a desire for dependency. Excessive drinking and/or speech difficulty may be present. There is a low level of drive or motivation.

DISCUSSION

An inspection of the profiles in Table 1 and Figure 1, and an examination of the impressions of the three "judges," indicate that the elevation of the *Hs*, *D*, and *Hy* are noteworthy. The elevation of these three scales, as well as the profile of the remaining scales, suggests the "neurotic triad" which the authors of the MMPI suggest as the characteristic of the "greater portion of these persons not under medical care who are commonly called neurotic, as well as individuals so abnormal as to need psychiatric attention." This type of curve suggests the "typical" personality configuration in which the person reacts to stress with depression and is further characterized by lack of self-confidence, tendency to worry, narrowness of interest, and introversion.

It is beyond the scope of this paper to review in detail, or to examine, the various concepts and theories concerning the etiology of multiple sclerosis or the "dynamics" of the personality structure. It is of interest to note that the present findings tend to support the common clinical observation of the presence of a "neurotic overlay" in multiple sclerosis. It is of further interest to note that according to the case histories of this veteran population of multiple sclerotics, approximately 65 per cent had received a diagnosis of some type of hysteria prior to the development of overt neurological symptoms which could be identified and a diagnosis of multiple sclerosis could be clearly established.

Grinker, Ham, and Robbins [2] have pointed out that neurotic manifestations often obscure the first evidence of the central lesions, causing many false diagnoses of hysteria for the entire syndrome. Grinker *et al.* indicate that these neurotic manifestations have been considered as a problem in differential diagnosis, and that only rarely has attention been given to the possibility that the disturbance may not only be the concomitant or result of the damage to the nervous system, but also an important factor in the etiology. In Sugar and Nadell's study [5], wherein special emphasis was placed on the emotional affective symptomatology, the changes in the emotional content or prevailing mood were considered an important diagnostic sign. They suggest that the affective disturbances in mood and behavior commonly noted are consistent with the patient's previous personality makeup. Although the present study is in agreement with their finding that disturbances in mood and behavior are of importance in understanding the multiple sclerotic patient, it is at variance with their findings insofar as it suggests the prevailing mood and the emotional content are in the direction of depression. Examination of personal diaries and other autobiographical material submitted by several patients in conjunction with the broader phases of the continuing work on multiple sclerosis, clearly shows that dysphoria is the predominant mood for a majority of this sample population of multiple sclerotics. One factor possibly explaining the difference in the present findings and those of Sugar and Nadell is the shorter average duration of illness in our patients. The patients in this study had been ill on an average of approximately four years at

the time of initial examination, whereas the patients in the Sugar and Nadell study had been ill for an average of 10.8 years at the time of examination.

The findings revealed by the MMPI profiles are not surprising in view of the impairment of functioning brought about by the disease process. Generally it has been observed that the patients have difficulty in accepting the disease and its progressive limitations. There is a frequently expressed attitude that physical strength and health are considered necessary for confidence and accomplishment, and that conflict between desires and ability to perform permeates the personality. There is, furthermore, an ambivalence concerning the future role in life, producing indecisiveness and feelings of insecurity which are often reflected in relatively poor emotional control and social adjustment. These psychological factors, of course, should be considered in any treatment program of multiple sclerosis.

The value of the present MMPI study, or any other psychometric measure, is not so much in obtaining proof of the obvious, as it is in establishing more objectively and precisely the nature and degree of the personality characteristics associated with the disease. A final note of caution to be kept in mind is that psychometric measures do not pretend to furnish a neurological diagnosis, nor do they determine whether or not a person is a multiple sclerotic.

SUMMARY

The personality characteristics associated with multiple sclerosis as revealed by the MMPI profiles are presented and discussed. The personality characteristics revealed are: a reaction of depression, preoccupation and concern about bodily functions, feelings of hopelessness and insecurity, as well as tendencies toward indecisiveness, narrowness of interests, and introversion. Difficulty in accepting the disease and its progressive limitations, as well as ambivalence and insecurity concerning the future, are often reflected in relatively poor emotional control and social adjustment.

Received September 6, 1950.

REFERENCES

1. CANTER, A. H. Direct and indirect measures of psychological deficit in multiple sclerosis. *J. gen. Psychol.*, 1951, **44**, 3-50.
2. GRINKER, R., HAM, G. C., AND ROBBINS, F. P. Some psychodynamic factors in multiple sclerosis. *Proc. 28th annu. Meet. Ass. Res. nerv. ment. Dis.*, December, 1948.
3. HATHAWAY, S. R., AND MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* Minneapolis: Univ. Minn. Press, 1943.
4. MEEHL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, **14**, 165-171.
5. SUGAR, C., AND NADELL, R. Mental symptoms in multiple sclerosis. *J. nerv. ment. Dis.*, 1948, **98**, 267-280.

AN INDEX OF ADJUSTMENT AND VALUES

ROBERT E. BILLS, EDGAR L. VANCE, AND ORISON S. McLEAN
UNIVERSITY OF KENTUCKY

SEVERAL authors have asserted that the basic human need is the drive to preserve and enhance self-organization. Lecky [8] emphasized the drive to maintain or enhance the consistency of the core of the personality, the value system. Snygg and Combs [12] prefer a different formulation and see as the fundamental need in behavior the maintenance and enhancement of the phenomenal self. Rogers [10] apparently includes physiological and psychological aspects in his basic drive and asserts the need to maintain or enhance self-organization. Stagner [13] holds that the self functions as a perceptual object for the individual, who, in turn, seeks to maintain perceptual constancy with regard to it. All of these views appear to have much in common.

Enhancement of psychological organization implies two characteristics: (1) that the individual has information relative to his present self-organization, and (2) that the individual has a view of himself as he wishes to be. The former been called the "self-concept" by Raimy [9] and the latter will be designated in this paper as the "concept of the ideal self."

From the point of view of phenomenological psychology, maladjustment may be defined as any discrepancy between the concept of self and the concept of the ideal self. For practical purposes it may be assumed that maladjustment exists when the discrepancy between the concept of self and the concept of the ideal self is great enough to cause unhappiness. For the sake of clarity it should be noted that this is a definition of *personal* maladjustment. *Social* maladjustment results from a conflict of the values of the individual with the values of his society.

The concept of self may be further defined as the traits and values which the individual has accepted as definitions of himself. Values are derived from traits, a trait being an adjective which may be used to describe a person. A *value* is a trait which the individual considers desirable.

The possession or acquisition of a particular value in the opinion of an individual makes him a better person. Being a particular kind of person, for example, being academic, is a value if possessing it causes the individual to feel he is a better person in his own eyes. Thus, traits are examined in light of our philosophy of life and become values or are rejected in accordance with their consistency with our value system or philosophy. The philosophy of life, the value system of the individual, and the concept of the ideal self are synonomous. The goal in life is to produce consistency within the value system. It is the striving for maintenance and enhancement of the consistency within the value system that motivates behavior.

Important in this system of thought are definitions of attitudes. An *attitude* is an evaluation. An individual may have attitudes toward traits, interests, and self. 1. An attitude toward a trait is a feeling or evaluation as to whether or not a certain trait constitutes a value. 2. An attitude toward self is a feeling or evaluation by the individual as to whether or not a trait which he possesses in a given amount constitutes a value. 3. An attitude toward an interest is a feeling as to whether or not this interest is helpful in maintaining or achieving a value or values. An interest is here considered to be a means or a path by which we achieve values. An interest is any behavior which is pursued by an individual with the evaluation that this activity will enable him to attain the values which he has accepted for himself.

DESCRIPTION OF THE INDEX

The Index of Adjustment and Values was designed to test the theoretical formulations given above, to serve as a research tool, and to assess changes in adjustment which occur during psychotherapy. In addition it may prove useful in revealing changes which occur within the individual's value system as a result of therapy by an examination of the individual items before and after therapy. In order to satisfy the requirements implicit in the rationale stated above and to assess changes in adjustment it is necessary to obtain measures of: (1) the self-concept of the individual; (2) the attitude which the individual holds toward himself in his present condition; and (3) the individual's concept of his ideal self, that is, the values toward which he is striving. The total of the discrepancies between the self-concept and the concept of the ideal self would be a measure of adjustment.

The purposes of the Index are strikingly similar to those basic to current investigations into personality theory which use Stephenson's Q Technique [14] and are now in progress at the Counseling Center of the University of Chicago [10, p. 140]. The Chicago investigations use a sample of a universe of self-referent statements obtained from recordings of client-centered therapy cases, whereas the Index is based on a sample of traits and values. It is quite possible that underlying the self-referent statements of the Chicago investigation are traits and values similar to those used in the Index. If this is true then the findings in the two lines of work should be similar.

To arrive at the Index a sample of 124 words was taken from Allport's list of 17,953 traits [2], In selecting this sample an effort was made to choose those items which occur frequently in client-centered interviews and which seem to present clear examples of self-concept definitions. It was hoped in this way to obtain an index which would show adequate test-retest reliability.

The 124 words which were selected were arranged in a vertical list and the words were followed by three blank columns. The subjects were asked to use each of the words to complete the sentence "I am a (an) ——— person" and to indicate on a five-point scale how much of the time this statement was like them. This rating was placed in the blank in Column I. The use of rating number one indicated a rating of seldom; number two, a rating of occasionally; number three, a rating of about half the time; number four, a rating of a good deal of the time; and number five, a rating of most of the time. Column I sampled, then, the concept of self.

The subjects were also instructed to indicate, in the second blank, a rating which would tell how they felt about themselves as described in the first blank. The ratings are as follows: (1) I very much dislike being as I am in this respect; (2) I dislike being as I am in this respect; (3) I neither dislike being as I am nor like being as I am in this respect; (4) I like being as I am in this respect; (5) I very much like being as I am in this respect. The sum of Column II measured acceptance of self.

The subjects were likewise instructed to use each of the words to complete the sentence "I would like to be a (an) ——— person" and to indicate in the third blank how much of the time they would like this trait to be characteristic of them. The same ratings were used as in the first blank. Thus, Column III sampled the concept of the ideal self.

The subjects were further instructed to complete the three ratings for each of the words before going to the next trait.

The list of 124 traits was given to 44 subjects and three weeks later it was given again to the same group. When the total number of ratings given by each student (376 on each test) is considered, it is doubtful that any part of the reliability of the test can be attributed to the memory of answers given on the first test. When the total acceptance of self scores (sum of Column II) on the first administration were correlated with the same scores on the second administration an r of .67 was obtained. Likewise a coefficient of .68 was obtained when a correlation of the discrepancy scores (difference between Column I and Column III summed without regard for sign) was made. Both of these coefficients are significantly different from zero at less than the .01 level of confidence.

The results of the above 44 tests were subjected to an item analysis to eliminate unreliable items. This analysis was accomplished by determining the difference between the ratings given in Column III on each test by each subject and totaling these differences to determine the total difference for each word and the total difference for each subject. Words which showed a greater variation than the variations of the subjects making the ratings were considered unreliable. This analysis resulted in the retention of the following 49 words: acceptable, accurate, alert, ambitious, annoying, busy, calm, charming, clever, competent, confident, considerate, cruel, democratic, dependable, economical, efficient, fearful, friendly, fashionable, helpful, intellectual, kind, logical, meddlesome, merry, mature, nervous, normal, optimistic, poised, purposeful, reasonable, reckless, responsible, sarcastic, sincere, stable, studious, successful, stubborn, tactful, teachable, useful, worthy, broadminded, businesslike, competitive, and faultfinding. The above is the order in which the traits appeared on the Index.

RELIABILITY OF THE INDEX

The revised index was given to 237 students at the University of Kentucky. When the acceptance of self scores (Column II of the Index) for each subject on the odd-numbered items were correlated with the acceptance of self scores on the even-numbered items a corrected r of .91 was obtained. This coefficient is significantly different from zero at less than the .001 level of confidence. The corrected split-half reliability of the discrepancy scores (difference between Column I and Column III summed without regard for sign) using odd-numbered versus even-numbered items was .88 for the same group. This coefficient is likewise significantly different from zero at less than the .001 level of confidence.

The Index was re-administered to 175 of the above 237 subjects six weeks after the first testing. The test-retest reliability of the acceptance of self scores for this group was .83 and for the discrepancy scores was .87. Both of these coefficients are significantly different from zero at less than the .001 level of confidence. These data show that the acceptance of self and discrepancy scores are reliable measures.

The acceptance of self and discrepancy scores of the above 175 subjects on their first test were correlated and a coefficient of —.77 was obtained. This coefficient is also significantly different from zero at less than the .001 level of confidence. These data show that those persons who score high on acceptance of self score low on discrepancy and those people with low acceptance of self scores show high discrepancy scores.

NORMS

The Index has been administered thus far to a total of 482 subjects, all college students. Scores for discrepancy and for acceptance of self are available for this group. The responses on each item in Columns I and III have been tabulated for 325 persons in this group. The distribution of acceptance of self and discrepancy scores approximate normal distributions. More cases are in the process of examination.

VALIDITY

The following investigations represent efforts to establish the validity of the Index.

Rorschach Results. Twenty female college students who were volunteers were first given the Index and then a Rorschach examination. The Rorschachs were divided into two groups on the basis of the presence of neurotic or psychotic signs as found in Klopfer [7]. This division placed 15 records in the first group and five in the latter. The acceptance of self scores of 14 of the 15 records having neurotic signs were below the mean of the 482 standardization subjects and all five of the records showing psychotic signs were above this mean. Thus, the mean of the acceptance of self scores appears to be dividing neurotic and psychotic groups. It should be emphasized, though, that these were experimental subjects and not therapy cases. One of the psychotic records came from a subject who had been hospitalized.

Two judges, working independently of each other, ranked the five records with psy-

chotic signs for level of adjustment. There was perfect agreement between the judges on the rankings. The subjects were also ranked in respect to their deviation from the mean of the acceptance of self scores. When these two ranks were correlated a *rho* of 1.00 was obtained. This *rho* is significantly different from zero at the .01 level of confidence.

One judge ranked the 15 Rorschach records with neurotic signs according to level of adjustment. These subjects were likewise ranked on the deviations of their acceptance of self scores from the mean of the acceptance of self scores given by the standardization group. This correlation gave a *rho* of .60, which is significantly different from zero at the .05 level of confidence.

It has been mentioned above that the acceptance of self scores given by the standardization group approximate a normal distribution. According to the Rorschach data, deviations below the mean of the standardization group are in the neurotic direction and deviations above the mean are in the psychotic direction. Normal scores seem to be centered around the mean of the distribution.

The Rorschach, thus, at least partially validates the acceptance of self score as a measure of adjustment. The Rorschachs also partially validate the mean acceptance of self score as an important dividing point.

Student-Centered Teaching. Several studies have shown that positive therapeutic effects follow the application of client-centered therapy methods to the classroom in the form of student-centered teaching techniques [1, 3, 4, 5, 6, 10, 11, 12]. It was predicted that the gains in adjustment resulting from these classes would be reflected in a change in scores on the Index. More specifically it was predicted that both the acceptance of self scores and the discrepancy scores would move a significant amount in the direction of the population mean for these scores.

Three classes in mental hygiene were tested at the beginning of the semester course and again at the end of the course.[1] The three classes included 38 students. As a result of the chance variations in scores it would be expected that at the .05 level of confidence one of 20 scores would vary more than 1.97 times the standard error of measurement of the Index calculated from the standard deviation of Column II and the test-retest reliability coefficient of Column II. Thus, it would be predicted that by chance alone two of the 38 students would show a change in acceptance of self scores greater than 1.97 times the standard error measurement. As a result of the test-retest of the three student-centered teaching classes it was determined that 14 of the students or seven times the expected number had made changes greater than the statistical criterion.

The conclusion that the Index measures the predicted changes and so constitutes a valid measure of these changes was thus strengthened, but a possible source of error lies in the length of the test-retest period for which the reliability was calculated. Data are being collected at the present time to test the reliability coefficient over a period of one semester. These data will serve as a source of controls for further investigation.

Threat from Self versus Threat from Others. If the sum of Column II of the Index measures, as it is purported to do, the acceptance of self then those subjects who rate themselves below the mean of the population are less accepting of self and those subjects who rate themselves above the population mean are more accepting of self. It follows, also, that subjects who score below the mean should blame themselves for their unhappiness in life since they feel they are essentially weak, whereas subjects who score above the mean would be ready to blame other people and factors outside themselves for their unhappiness.

Students in two classes were told that "all people are at least somewhat unhappy. This does not necessarily mean that you are an unhappy person but that you may not be completely happy. Will you please briefly itemize the reasons for your unhappiness?" These students had been tested with the Index one week prior to this exercise and at that time they were assigned numbers to keep their in-

[1]The writers wish to express appreciation to Professor Paul Bowman of the University of Louisville for gathering data from two of his classes for this study.

dentities anonymous. The students were asked to place the same number on the list of reasons for unhappiness. Data were thus available on this question from 142 students.

The statements which the students made were scored independently by two judges for the direction of perceived threat. On one class the judges were in agreement in 75 per cent of the cases and on the other class the judges agreed in 74 per cent of the cases. In the cases in which the judges disagreed, agreement was arrived at in conference. A dichotomous chi-square was computed using for one dichotomy scores above and below the mean for acceptance of self and for the other dichotomy threat from self versus threat from outside self. The calculated chi-square was 24.6. This chi-square may be interpreted to mean that at less than the .001 level of confidence the distribution was significantly different from chance. Acceptance of self scores below the population mean were significantly related to threat from self and acceptance of self scores above the population mean were significantly related to threat from others.

CONCLUSION

The data show that the Index of Adjustment and Values is a reliable instrument which should prove useful as a research tool. Further study of the reliability of the total scores and the individual items appears necessary. Data which are now being collected should be useful in this problem.

The data which have been collected indicate that the Index is valid. The results thus far have been encouraging, and further validation studies are now in progress.

SUMMARY

By use of 49 traits an Index of Adjustment and Values has been designed which appears to measure the values of a person, his acceptance of self, and the discrepancy which exists between his concept of self and his concept of his ideal self. Corrected split-half reliability coefficients of .91 and .88 were obtained for a group of 237 students. Test-retest reliability coefficients over a period of six weeks for a group of 175 students were .83 and .87. Three studies indicate that the Index is a valid measure of adjustment and values.

Received March 9, 1951.
Early publication.

REFERENCES

1. ALBRECHT, M., AND GROSS, L. Nondirective teaching. *Sociol. soc. Res.*, 1948, **32**, 874-888.
2. ALLPORT, G. W., AND ODBERT, H. S. Trait-names: A psycho-lexical study. *Psychol. Monogr.*, 1936, No. 211.
3. AXLINE, VIRGINIA M. *Play therapy.* Boston: Houghton Mifflin, 1947.
4. AXLINE, VIRGINIA M. Nondirective therapy for poor readers. *J. consult. Psychol.*, 1947, **11**, 61-69.
5. FAW, V. A. A psychotherapeutic method of teaching psychology. *Amer. Psychologist*, 1949, **4**, 104-109.
6. GROSS, L. An experimental study of the validity of the nondirective method of teaching. *J. Psychol.*, 1948, **26**, 243-248.
7. KLOPFER, B., AND KELLEY, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1946.
8. LECKY, P. *Self-consistency.* New York: Island Press, 1945.
9. RAIMY, V. C. Self reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153-163.
10. ROGERS, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
11. SHEDLIN, A. J. A student centered class. *Personal Counselor*, 1947, **2**, 116-131.
12. SNYGG, D., AND COMBS, A. W. *Individual behavior.* New York: Harper, 1949.
13. STAGNER, R. Homeostasis as a unifying concept in personality theory. *Psychol. Rev.*, 1951, **58**, 5-17.
14. STEPHENSON, W. A statistical approach to typology: The study of trait universes. *J. clin. Psychol.*, 1950, **6**, 26-37.

ITEM ANALYSIS AID FOR THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

GEORGE C. CLARK AND ROBERT M. ALLEN
THE UNIVERSITY OF MIAMI

ITEM analysis of the group form of the Minnesota Multiphasic Personality Inventory [1] can provide a wealth of clinical information. This is evident from the fact that the testee must answer 566 statements in nine psychiatrically defined areas.

The present profile or quantitative method of interpretation can show only how *much* the testee has deviated rather than *how* he deviated from the normal limits. It has been found that profile charts showing little or no deviation conceal much that is valuable. Or, the presence of a peak on a profile chart may indicate the need for further testing, but without item analysis much of the significant information is overlooked.

The Item Analysis Aid (IAA) devised by the authors permits a rapid method of individual item analysis during the scoring period. Each scale may be analyzed separately, including *L*, *F*, and *K*. The IAA consists of a booklet containing the Inventory statements arranged according to their appearance in the respective manual scoring-key cards. Each statement in this IAA booklet has either a plus (+) or minus (—) sign in front of it, with the exception of the *L*-scale statements. A plus (+) sign indicates that the statement is significant for the particular scale if answered "True" by the testee. Similarly, a minus (—) sign means that the item is critical for the scale if marked "False" by the testee on the answer sheet. This booklet is bound so that it may be placed to the right of the IBM answer sheet for easier transition between scoring key and items in the booklet.

In addition, a second booklet has been prepared showing the manual scoring-key cards with circles representing the holes for the critical items (that are totaled to give the raw score for each scale). Each hole has a number which corresponds to the item (for that key) in the IAA booklet. This number may be written directly on the regular scoring-key cards purchased from the publisher [1]. Thus this second booklet may be discarded after transferring the numbers. The tester may go from the scoring-key card to the IAA booklet and read the critical statement, making a note of its content and the subject's response to it.

This supplementary aid was devised for the sole purpose of helping the MMPI user to increase its informational value.[1]

Received February 19, 1951.
Early publication.

REFERENCES

1. HATHAWAY, S. R., AND MCKINLEY J. C. *The Minnesota Multiphasic Personality Inventory.* (Rev. Ed.) New York: Psychological Corp., 1943.

[1]The authors have several mimeographed copies of the IAA booklets available and will send them to interested clinical psychologists while the supply lasts. Please address the senior author.

NEW BOOKS AND TESTS

Books

BABCOCK, HARRIET. *The MacQuarrie Test as a clinical instrument.* Lancaster, Pa.: Science Press, 1950. Pp. vii + 72. $2.00.

This monograph presents evidence for a new measure of efficiency of mental functioning, obtained by comparing a vocabulary age with corresponding age scores obtained from the MacQuarrie Test. The latter test is certainly as suitable as any for measuring a deteriorating mental function because of its dependence on speed, skill, and spatial perception. Although the monograph does not establish the validity of the Vocabulary-MacQuarrie difference beyond all doubt, there is considerable confirming evidence of a clinical nature. Norms are based on 1000 cases. No correction for age is provided, and the author argues against its use.

BERGMAN, MOE. *The audiology clinic.* Stockholm, Sweden: Acta Oto-Laryngologica, Suppl. 89, and 1104 S. Wabash Ave., Chicago: The Audiology Foundation, 1950. Pp. 107. $1.00.

Although published in Sweden, this is a thoroughly American monograph describing the construction, instrumentation, and operation of the clinic for the rehabilitation of the acoustically handicapped at the New York Regional Office, Veterans Administration. Many aspects of the monograph — the plans for soundproof rooms, the descriptions of equipment, the therapeutically-oriented rehabilitation program — will be of wide interest to clinical psychologists.

DEVEREUX, GEORGE. *Reality and dream, psychotherapy of a plains Indian.* New York: International Universities Press, 1951. Pp. xxv + 450. $7.50.

Here is an outstanding example of idiographic research in which the integrated application of anthropology, psychoanalysis, and psychology to an individual subject has enriched our knowledge of man, of culture, and of the psychotherapeutic process. "Jimmy Picard," a war-veteran plains Indian with a history of an ill-defined head injury, was admitted to a VA hospital as possibly psychotic. The anthropological-therapeutic work reveals Jimmy as having a moderate chronic anxiety, evoked by situational stress, and manifested in ways understandable only in terms of his culture and his individuality. The volume contains near-verbatim reports of thirty interviews, and protocols of psychological tests.

DIMICHAEL, SALVATORE G. (Ed.) *Vocational rehabilitation of the mentally retarded.* Federal Security Agency, Office of Vocational Rehabilitation, Rehabilitation Service Series No. 123. Washington: U.S. Government Printing Office, 1950. Pp. vii + 184. 45¢.

The modern attitude toward the mentally retarded is concerned mainly with their training and their readjustment in society, a far cry from the misconception of a generation ago which tended to regard the deficient as entirely a liability. This able manual, the product of the editor and twelve collaborators, gives some background on the problem of retardation and devotes most of its pages to counseling, education, family and community relations, and employment. It should find a place in training curricula and on the reference shelves of psychologists.

GARRISON, KARL C. *Psychology of adolescence.* (4th Ed.) New York: Prentice-Hall, 1951. Pp. xxiii + 510. $6.00.

The third revision of this standard text since its first publication in 1934 gives added material on physiological changes, peer relations, attitudes and beliefs, and moral and spiritual development. The approach is eclectic and research-oriented. The considerable number of references dated subsequent to the last revision gives the impression that the reworking of the material has been thorough.

LINDQUIST, E. F. (Ed.) *Educational measurement.* Washington: American Council on Education, 1951. Pp. xix + 819. $6.00.

A comprehensive reference work and graduate textbook on the entire field of educational measurement, the product of twenty authors and fifty-one

NOTE: Some reviews in this issue were prepared by the Associate Editors, who may be identified by their initials. Unsigned reviews are by the Editor. —*L. F. S.*

collaborators working over a four-year period. The volume's three main parts are on the functions of measurement, the construction of tests, and measurement theory. Clinical psychologists have much to gain from the study of such a book, for no gulf exists between educational measurement and psychological measurement, and the same techniques apply to both. Technical handicaps from which many clinical tests suffer could be lessened by improved methodological sophistication. There is no reason to believe that an instrument would lose any of its clinical power by also being a good test.

LOEWENSTEIN, RUDOLPH M. *Christians and Jews, a psychoanalytic study.* (Trans. by Vera Damman.) New York: International Universities Press, 1951. Pp. 224. $3.25.

Although called a psychoanalytic study, Loewenstein's examination of the relationships between Christians and Jews draws from many sources, from historical and cultural roots as well as from psychological investigations of individuals. The result is a reasonable and dispassionate account that carries conviction because of its very simplicity. The reader is likely to feel that he has always known and believed what the book presents, but such a reaction is more a tribute to the author's skill than a criticism of his lack of originality.

LORAND, SANDOR. (Ed.) *The yearbook of psychoanalysis.* New York: International Universities Press, 1950. Vol. 6. Pp. 307. $7.50.

This is the sixth annual *Yearbook of Psychoanalysis.* Like its predecessors it is a collection of reprints of important psychoanalytical journal articles intended to represent a year's progress in psychoanalysis. Its twenty contributions range from Bernfeld on "Freud's Scientific Beginnings," to Eisenstein's "Dreams Following Intercourse." The range of topics is broad, the list of contributors imposing. A two-page selected reading list is appended. It is excellent reading for the busy clinician who wishes to keep abreast of the psychoanalytic literature.—*W. A. H.*

MOWRER, O. HOBART. *Learning theory and personality dynamics.* New York: Ronald, 1950. Pp. xviii + 776. $7.50.

This volume is unique in that it is a book of readings of theoretical, experimental, and applied publications by a single author in what were regarded a few years ago as two somewhat distinct fields. The book is one of several recent publications which attest to the happy and productive marriage of personality and learning theory. Several of the reprinted articles have frequently been quoted and assigned as required readings for advanced students in learning. Others are highly important contributions to personality theory. The range of articles may be illustrated by citing some chapter titles: "Avoidance Conditioning and Signal Duration — A Study of Secondary Motivation and Reward," "The Life and Work of Edgar Allan Poe — A Study in Conscience-Killing," "On the Psychology of 'Talking Birds' — A Contribution to Language and Personality Theory." The articles are presented under two major headings, learning theory and personality dynamics. Some academic psychologists will no doubt find interest in only one group of articles, but the integration of animal experimentation and personality study is recognized as important in the training of clinical psychologists, among others. It may be well for some psychologists with limited interests to be made aware of this broadening in the scope of contemporary psychology. The diversity of the articles presents a versatility seldom found among psychological specialists of the high competence of Dr. Mowrer. It is heartening to see one who, with undiminished scholarship, is capable of designing crucial experimental studies applying these insights to practical problems. Certain of these papers are basic, while others represent the application of the author's hypotheses to more particular and complex problems. Some of the previously unpublished articles might have been omitted without weakening the central integrating theme of the volume. The papers vary in technical level and are intended for different audiences. Despite this heterogeneity, the book is integrated by the author's notes, his prefatory remarks at the beginning of each chapter, and cross-references throughout. The material rests on a bibliography requiring 20 pages.—*F. McK.*

PORTER, E. H., JR. *An introduction to therapeutic counseling.* Boston: Houghton Mifflin, 1950. Pp. xi + 223. $3.00.

As Rogers says in his introduction, "This is not an easy book to classify." It is essentially a stimulus to a series of experiences, not *about* counseling but *of* counseling, through exercises based on verbatim extracts from counseling interviews. The central topics include self-evaluative attitudes, responding to the internal frame of reference, problems of beginning, middle, and closing phases, and an introduction to interview analysis. There is a pretest to evaluate the attitudes and beliefs that the learner brings and a posttest to indicate their change. The author's attitude is frankly nondirective, but he persuades by example rather than by preaching. The book implements the concept of student-centered teaching in an area to which it is most applicable, and will prove interesting and challenging to all teachers of therapy, regardless of their theoretical orientations.

RAIMY, VICTOR C. (Ed.) *Training in clinical psychology.* New York: Prentice-Hall, 1950. Pp. xix + 253. $4.00.

This is a systematized compilation of deliberations by the staff of the Conference on Graduate Education in Clinical Psychology held at Boulder,

Colorado, in August, 1949, and sponsored by the American Psychological Association under grant from the National Institute of Mental Health. E. Lowell Kelly, Chairman of the Executive Committee which organized the conference, acknowledges indebtedness to Raimy for preparation of the report. The 71 participants were, with only a few exceptions, representative of teaching and administrative personnel from university and government agencies. The trend of the discussion appears to reflect a major emphasis toward VA and PHS identification of interests. Awareness of the needs for both professional and scientific training is recognized as "something of an educational experiment" (p. v), and accepted as the apparent academic point of departure. This reviewer would have preferred the title *Education for* rather than *Training in* Clinical Psychology as a sounder orientation. He would also like to have seen included a more definite and comprehensive elucidation of both general and specific fields of clinical practice and research, as well as a clearer delineation of functions and procedures as the ultimate criteria of the proposed training. He regrets that the conference did not more specifically follow the leads so tactfully presented in the introductory paper by Felix, that little attention was paid to future possibilities and present necessities, that the Ph.D. requirement continues as a symbol rather than a guarantee of clinical competence, that in spite of a rather pious concern for protecting the public there is so little emphasis on a warm devotion to the needs of people of all kinds and ages who are in deep trouble and who go to other ministers for aid and comfort. The conference evidently represented a sincere and earnest effort. But is the teaching of clinical psychology going to promote or merely capitalize the needs considered? In the former case the problems of practice are paramount and must receive field representation. The Ph.D. requirement may dignify research but appears unlikely, as inferred from the report of this conference, to assure competence for the varied requirements of practice. The report impresses this reviewer as a masterpiece of collation and a chef-d'oeuvre of publication.—***E. A. D.***

ROGERS, CARL R. *Client-centered therapy.* (With chapters by Elaine Dorfman, Thomas Gordon, and Nicholas Hobbs.) Boston: Houghton Mifflin, 1951. Pp. xiii + 560. $4.00.

Although it is not quite nine years since the publication of *Counseling and Psychotherapy,* developments in nondirective or client-centered counseling have been so extensive that Rogers has evidently felt impelled to bring his earlier publication up to date; to describe recorded experience with the approach; to present research results; and to show the extension of the method to such fields as play therapy, group therapy, group leadership, teaching, and the training of counselors. Not the least important section of the book presents tentatively "a theory of personality and behavior," based upon the author's experience with nondirective counseling. The nineteen propositions comprising this theory seem rather familiar, probably because they are an admixture of Gestalt psychology, aspects of Freudian psychology, and original contributions of Rogers, particularly those relating to concept of self, and the organization of experience. The merit of this theory of personality lies in its clinical meaning and in the organization of clinical experience which Rogers has injected into previously generalized verbalizations. *Client-Centered Therapy* goes a long way toward lessening the gap between nondirective and other types of therapy, particularly in its treatment of transference and diagnosis. Although not intended to replace *Counseling and Psychotherapy,* it undoubtedly will. For some years to come, it will probably be a standard work in the counseling field.—*M. K.*

SCHILDER, PAUL. *Brain and personality.* (2nd Printing). New York: International Universities Press, 1951. Pp. v + 136. $2.50.

A reissue of Schilder's monograph, which first appeared in *Nerv. ment. Dis. Monogr.* in 1931. It consists of two series of lectures, on psychological aspects of cerebral neuropathology, and on the personality and motility of schizophrenics.

STOLZENBERG, JACOB. *Psychosomatics and suggestion therapy in dentistry.* New York: Philosophical Library, 1950. Pp. xi + 152. $3.75.

A dentist writes for dentists on a number of psychological topics: neurotic defenses in dental patients, psychosomatic syndromes encountered in dentistry, suggestion therapy, and the use of hypnosis for the relief of dental operative pain. It is heartening to find some dentists now looking at the whole person instead of only at his teeth, but the superficial and symptom-oriented approach seems to repeat errors that psychiatry overcame a generation ago.

WEPMAN, JOSEPH M. *Recovery from aphasia.* New York: Ronald Press, 1951. Pp. xii + 276. $4.50.

More and more clinical psychologists are concerned with the effects of gross brain damage, either directly or as members of a rehabilitation team with speech pathologists, neurologists, educators, and others. To meet the need for training the participants in such a team, Wepman has produced a textbook on aphasia and its treatment which is so well planned and written that its clarity may be mistaken for elementariness. The book views aphasia as a disorder of the whole person, and is as much interested in the self-concept of the aphasic as in an analysis of his impaired and residual communication functions. The latter are not neglected, however. The book contains an adequate introduc-

tion to theories of aphasia, a survey of recent research, and an extended discussion of treatment.

YACORZYNSKI, G. K. *Medical psychology.* New York: Ronald Press, 1951. Pp. xviii + 535. $6.00.

A selective general psychology, emphasizing the areas needed for preparation for psychiatry and clinical psychology. The three main parts of the book take up psychobiological principles (needs, emotions, learning, perception, adjustment), inheritance and maturation, and the structuralization of personality. The author is quite successful in his avowed attempt to integrate learning-theory, Gestalt, and psychoanalytic approaches. Beginning students will find the book smoothly organized and free from conflicts and will hardly realize how well they have been prepared for the more advanced study of all of the main current approaches to behavior and personality. Incidentally, first-year graduate students will find the book useful for a clinically-oriented review of general psychology.

Epidemiology of mental disorder. Milbank Memorial Fund. New York: Milbank Memorial Fund, 40 Wall St., 1950. Pp. 198. 50¢.

This is a report of the papers and discussion at the 1949 Annual Conference of the Milbank Memorial Fund. Thirty-eight persons participated, including one psychologist, one anthropologist, several sociologists and medical statisticians, and the remainder, psychiatrists and epidemiologists. The conference was called to explore the possibility of epidemiological studies of mental disorders, and a good deal of time was devoted to ironing out differences in terminology and in types of approach. Some data are presented in the papers, but the volume is chiefly valuable for the content of the discussions, which cannot be easily summarized, and for the 362-item bibliography.—*A. R.*

Twentieth anniversary review of the Josiah Macy, Jr., Foundation. 565 Park Ave., New York: The Josiah Macy, Jr., Foundation, 1950. Pp. 110.

The report traces the contributions of the first twenty years of the Foundation, in which over three and one-half million dollars have been expended for planning conferences and for research on growth and development, aging, psychiatry, medical education, and interprofessional relationships.

CORRECTION

Due to a printer's error, the second book review in the righthand column of page 169 of the April, 1951, issue of this Journal appeared with an incorrect heading. The review, with its correct caption, is repeated as follows:

MIKESELL, W. H. (Ed.) *Modern abnormal psychology.* New York: Philosophical Library, 1950. Pp. ix + 880. $10.00.

Twenty-four chapters, each written by a specialist, are "designed to meet the interest of the general public in abnormal psychology, as well as the requirements of the student for a reference and source book." Bibliographical references accompany each chapter. The coverage is good, style and exposition are adapted to the intended audience, and the treatment is adequate at the desired level. It should be useful for its purpose.—*W. A. H.*

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Volume 64, 1950

Patterns of Personality Rigidity and Some of Their Determinants. Seymour Fisher Elgin State Hospital. $307, $1.00

The Value of an Oral Reading Test for Diagnosis of the Reading Difficulties of College Freshmen of Low Academic Performance. Charles A. Wells, American International College. $308, $1.00

Rorschach Responses Related to Vocational Interests and Job Satisfaction. Solis L. Kates, Michigan State College. $309, $1.00

Symbol Elaboration Test (S.E.T.): The Reliability and Validity of a New Projective Technique. Johanna Krout, Chicago Psychological Institute. $310, $2.00

Changes in Responses to the Minnesota Multiphasic Inventory Following Certain Therapies. William Schofield, University of Minnesota. $311, $1.00

A Scale for Measuring Teacher-Pupil Attitudes and Teacher-Pupil Rapport. Carroll H. Leeds, Furman University. $312, $1.00

The Nature and Efficacy of Methods of Attack on Reasoning Problems. Benjamin Burack, Roosevelt College. $313, $1.00

The Validity of a Multiple-Choice Projective Test in Psychopathological Screening. Martin Singer, VA Hospital, Long Island. $314, $1.00

A Normative Study of the Thematic Apperception Test. Leonard D. Eron, Yale University. $315, $1.50

Experimentally Induced Variations in Rorschach Performance. Edith E. Lord, Arizona State Department of Health. $316, $1.00

An Evaluation of Personality-Trait Ratings Obtained by Unstructured Assessment Interviews. Ernest C. Tupes, United States Air Force. $317, $1.00

American Psychological Association

1515 Massachusetts Avenue N.W., Washington 5, D.C.

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Volume 63, 1949

Facial Expressions of Emotion. James C. Coleman, University of Southern California. $296, $1.00

A Comparative Study of the Wherry-Doolittle and a Multiple Cutting-Score Method. Glen Grimsley, General Motors Institute. $297, $.75

Factor Analyses of Tests and Criteria: A Comparative Study of Two AAF Pilot Populations. William B. Michael, Princeton University. $298, $1.00

The Appraisal of Parent Behavior. Alfred L. Baldwin, Joan Kalhorn, and Fay Huffman Breese, Fels Research Institute. $299, $1.50

Studies of Identical Twins Reared Apart. The late Barbara S. Burks; and Anne Roe, New York City. $300, $1.00

Color Preferences of Psychiatric Groups. Samuel J. Warner, New York City. $301, $.75

Perception of Body Position and of the Position of the Visual Field. H. A. Witkin, Brooklyn College. $302, $1.00

An Experimental Examination of the Thematic Apperception Technique in Clinical Diagnosis. A. A. Hartman, Boston University. $303, $1.00

Religion and Humanitarianism: A Study of Institutional Implications. Clifford Kirkpatrick, Indiana University. $304, $.75

The Development and Validation of a Set of Musical Ability Tests. Robert W. Lundin, Hamilton College. $305, $1.00

A Comparison of Two Tests of Intelligence Administered to Adults. Anna S. Elonen, University of Chicago. $306, $1.00

The 1949 volume of the Psychological Monographs consists of eleven separate issues. Orders for any of these can be placed separately at the prices listed above, or the entire volume can be ordered for $6.00.

American Psychological Association

1515 Massachusetts Avenue N.W., Washington 5, D.C.